Trajectories of work absence in England due to a musculoskeletal or mental health condition: an electronic health record study

Amardeep Singh Legha

A thesis submitted for the degree of
Doctor of Philosophy

October 2024

Keele University

Declaration

This thesis is based in part on data from the Clinical Practice Research Datalink obtained under licence from the UK Medicines and Healthcare products Regulatory Agency. The data is provided by patients and collected by the NHS as part of their care and support. The interpretation and conclusions contained in this study are those of the author/s alone.

Abstract

Introduction

Long-term sickness absence has recently been rising in the UK, causing a record number of economically inactive individuals and prompting concern from the Government. The aim of this thesis is to derive trajectories of absence over time due to a musculoskeletal (MSK) or mental health (MH) condition in an English population and explore association of these absence trajectories with health and sociodemographic characteristics.

Methods

A national primary care dataset, the Clinical Practice Research Datalink Aurum, was used to uncover trajectories of absence through issuance of fit notes. Different trajectory derivation methods were tested based on latent class analysis, alongside different approaches to specifying time intervals and follow-up periods. Trajectory-covariate association analysis was performed through multivariable multinomial logistic regression.

Results

The optimal chosen models contained n=43,130 and n=62,355 individuals with an incident fit note due to a MSK or MH condition, respectively. Five common trajectories were uncovered for both the MSK and MH condition fit note cohorts, using latent class growth analysis and based on two-monthly intervals over a one-year follow-up post index fit note. The two most common trajectories consisted of low absence (a ‘Single’ fit note and ‘Short Term’ absence), whilst the two least common trajectories were characterised by longer-term absence (‘Chronic Sustained’ and ‘Chronic Fast

Decreasing'), and the fifth by intermittent absence. Individuals associated with the most severe absence trajectories were: older, living in the North or Midlands or most deprived areas of England, prescribed opioids, and current smokers.

Conclusions

This thesis has highlighted different patterns of sickness absence due to a MSK or MH condition and profiles of individuals associated with intermittent and longer-term absence. Earlier and more targeted health and work intervention towards these high-risk subgroups, alongside policy interventions to reduce health inequalities, could help alleviate Britain's missing worker problem.

Acknowledgments

Firstly, I would like to express the utmost gratitude towards my supervisors: Professor Gwenllian Wynne-Jones, Professor Kelvin Jordan, and Professor Clare Holdsworth. Thank you for your guidance and support along this journey, and especially for remaining patient with me during this past year and maintaining an unwavering sense of calm in all of our supervision meetings. This thesis would not have been possible without your support. In particular, thank you to my lead supervisor, Professor Gwenllian Wynne-Jones, for providing timely feedback, constant encouragement, staying behind for all those extra hours when our meetings over-ran, and for helping me to become a better overall researcher.

I would also like to thank the Economic and Social Research Council and North West Social Science Doctoral Training Partnership for funding my research and the many training opportunities provided, and the School of Medicine for the opportunity to undertake this PhD in a conducive and friendly environment.

I have really appreciated the support and friendship of other students in the School, particularly during the COVID-19 pandemic – a sincere thanks. Many thanks to the staff in the School too for their support, in particular Dr Victoria Welsh for your enthusiasm and detailed feedback on all areas of my research, and Dr Kayleigh Mason and James Bailey for helpful advice in navigating the Clinical Practice Research Datalink datasets.

Finally, thank you to my friends and family who have helped keep me sane during this PhD and enjoy this final stage of student life.

Contents

List of Tables

List of Figures

List of Boxes

Abbreviations

AIC = Akaike Information Criterion

BAME = Black, Asian and Minority Ethnic

BIC = Bayesian Information Criterion

BLRT = Bootstrapped LRT

BMI = Body Mass Index

CAC = Central Advisory Committee

CASP = Critical Skills Appraisal Program

CCI = Charlson Comorbidity Index

CFI = Comparative Fit Index

CI = Confidence Interval

CiPCA = Consultations in Primary Care Archive

CMD = Common Mental Disorder

COPD = Chronic Obstructive Pulmonary Disease

CPRD = Clinical Practice Research Datalink

CSDH = Commission on Social Determinants of Health

DHSC = Department of Health and Social Care

DOI = Digital Object Identifier

DWP = Department for Work and Pensions

EHR = Electronic Health Records

EMIS = Egton Medical Information Systems

eRAP = electronic Research Applications Portal

ERC = Expert Review Committee

ESA = Employment and Support Allowance

FIML = Full Information Maximum Likelihood

GCM = Growth Curve Model

GEE = Generalized Estimating Equations

GFC = Global Financial Crisis

GMM = Growth Mixture Modelling

GMM-CI = GMM Class Invariant

GMM-CV = GMM Class Variant

GP = General Practitioner

GPRD = General Practice Research Database

GRoLTS = Guidelines for Reporting on Latent Trajectory Studies

HCP = Healthcare Professional

HMO = Health Maintenance Organisations

HWLE = Healthy Working Life Expectancy

IC = Information Criteria

ICD-10 = 10th revision of the International Statistical Classification of Diseases and Related Health Problems

ICSs = Integrated Care Systems

IMD = Index of Multiple Deprivation

IMRD = IQVIA Medical Research Data

IRR = Incidence Rate Ratio

ISAC = Independent Scientific Advisory Committee

LCA = Latent Class Analysis

LCGA = Latent Class Growth Analysis

LGCM = Latent Growth Curve Modelling

LMR-LRT = Lo-Mendell-Rubin LRT

LRT = Likelihood Ratio Test

LRV = Latent Response Variable

LSOA = Lower-Layer Super Output Area

LTA = Latent Transition Analysis

MAR = Missing at Random

MeSH = Medical Subject Headings

MH = Mental Health

MHRA = Medicines and Healthcare products Regulatory Agency

MSK = Musculoskeletal

NHS = National Health Service

NICE = National Institute for Health and Care Excellence

NIHR = National Institute for Health Research

NOS = Newcastle-Ottawa Scale

ONS = Office for National Statistics

OR = Odds Ratio

PPIE = Patient and Public Involvement and Engagement

PRISMA-P = Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols

QUIPS = Quality In Prognosis Studies

RDG = Research Data Governance

RMSEA = Root Mean Square Error of Approximation

RTW = Return-To-Work

SAIL = The Secure Anonymised Information Linkage

SEM = Structural Equation Modelling

SPA = State Pension Age

SRHIE = Strategic Review of Health Inequalities in England

SRMR = Standardised Root Mean Residual

SSP = Statutory Sick Pay

THIN = The Health Improvement Network

TLI = Tucker-Lewis Index

VAMP = Value Added Medical Products

WHO = World Health Organization

WRMSD = Work Related Musculoskeletal Disorder

Chapter 1. Background

1.1 Overview of Long-Term Sickness Absence

Long-term sickness absence, the focus of this thesis, has recently been rising in the UK and has become a topical issue. A report from the Office for National Statistics (ONS) showed that from spring 2019 to summer 2022, the number of working-age adults that were economically inactive (not working nor seeking employment) due to long-term sickness absence had increased by half a million.[1]

The UK Government also expressed concern in this increase in economic inactivity through a report entitled 'Where have all the workers gone?' (2022),[2] and placed particular emphasis on the detrimental consequences to the national economy. A loss of workforce can lead to increased inflation, as employers are obligated to increase wages to encourage staff recruitment, which leads to an increase in the cost of producing goods and services.[2] Furthermore, less workers results in less goods and services being produced, and at a higher cost due to the increased inflation, which restricts the overall growth of the economy. Finally, another avenue of concern with a shrinking workforce is the decline in contributions towards national taxes, which limits the availability of Governmental funds for public expenditure, and this is further exacerbated by increased welfare costs for those on long-term sickness absence.[2]

Thus, research that aims to reduce long-term sickness absence can be important for a healthier economy.

Aside from sickness absence affecting employers and wider society, there is also an impact on the individuals themselves to consider. Being in work is generally beneficial for physical and mental health, whilst being off work is associated with poorer health

and well-being.[3] For example, being part of the workforce can provide employees with a sense of purpose and social inclusion, which promotes better mental health.[4]

There are also financial consequences for the individual, as sick pay, paid by employers as Statutory Sick Pay (SSP) in the UK, entitles eligible claimants to £109.40 per week for up to twenty-eight weeks.[5] This is considerably lower than the UK weekly median earnings of £682, as reported by the ONS for April 2023.[6]

Additionally, another important concern is that being without employment due to sickness absence is a known driver of widening health inequalities,[7] which has implications regarding social justice and fairness. Wilkinson and Pickett argued that more equal societies experience less problems than more unequal societies, and used data collected over 30 years to show that outcomes relating to health and social problems such as physical and mental health, drug abuse, education, and social well-being were worse in countries with more inequality, irrespective of the overall wealth of the country.[8] They emphasized that within country differences were stronger drivers for poor social and health outcomes, not between country differences.[8]

Whilst the majority of people with a sickness absence tend to return-to-work (RTW) quickly, approximately 10% go on to have longer-term absences of >12 months.[9] The problem is that once an individual is on a long-term absence, it becomes progressively harder to RTW, and more adverse health and social problems are experienced.[3] As shown through the overall number of missing workers, this subgroup of long-term sick individuals can present significant issues for the overall UK economy.

Research that aims to identify individuals at risk of a long-term sick absence, at an early stage of their absence, and to subsequently help such individuals RTW quicker (for

example, through earlier and more targeted intervention), could have positive implications for the individual, their employers, and wider society.

In the remainder of this introduction Chapter, the theoretical and societal context for this thesis is presented. To begin, health inequality is defined, and several examples are provided to illustrate the impact of health inequality.

1.2 Health Inequalities

1.2.1 Health Inequality and Its Impact

Health inequality is defined by the National Institute for Health and Care Excellence (NICE) as "differences in health across the population, and between different groups in society, that are systematic, unfair and avoidable".[10] To reduce health inequality is one of the core principles of NICE.

The groups that experience health inequalities can, for example, be defined by socioeconomic status, deprivation, age, disability, sex, ethnicity, and geographical region, amongst others.[11] Then, the unjust differences in health between such groups can relate to health outcomes concerning not only the length of life, but also the health outcomes that affect the quality of life (e.g. prevalence of morbidity).

The consequences of health inequalities are considerable. An example of a report performed to evaluate the state of health inequities at a global level, was the World Health Organization (WHO) Commission on Social Determinants of Health (CSDH) report that was conducted in 2008.[12] This report concluded that "social injustice is killing people on a grand scale",[12(p40)] and noted that differences in where a child is born can significantly alter their life expectancy. For example, the authors showed that life expectancy was more than 80 years in countries like Japan or Sweden, down to around

72 years in Brazil, further down to 63 years in India, and in several African countries less than 50 years. Other factors influencing life expectancy included: socioeconomic position, ethnicity, type of work (temporary vs permanent), and education.

It is not only globally that marked health inequities exist, but also locally between the societies within a particular country or a group of countries, as illustrated for the UK in the next Section.

1.2.2 Early Development of Health Inequality Research in the UK

The report that was fundamental in bringing health inequality to the forefront of public policy and academic research in the UK was The Black Report (otherwise known as 'the Report of the Working Group on Inequalities in Health'), published in 1980.[13]

The Black Report studied the relationship between ill health and mortality in England and Wales, demonstrating that since the introduction of the National Health Service (NHS) in 1947 (which continues to provide healthcare free at the point of contact to UK citizens today), differences in risk of mortality had not reduced between different social classes, rather, these health inequalities had actually increased from the 1950s to the 1970s.[14] The foremost finding from this report was that this was due to various social determinants of health inequalities rather than the NHS per se, for example, relating to differences in: income, nutrition, education, and housing.[14]

To reduce these health inequalities, the authors of the Black Report suggested a range of social policy interventions for the Government to consider implementing, yet when the report was published there had since been a change of Government, and there was no longer support for these recommendations due to the associated financial expenditure.[14]

Nonetheless, following this groundbreaking report, there was an ensuing succession of research focusing on social inequalities that influence health.

For example, a major longitudinal study which is still ongoing today, the Whitehall II study investigated health inequalities amongst 10,314 British civil servants between 1985 and 1988.[15] Extending their work from the original Whitehall study that began in 1967,[16] their findings continued to show that social class (defined as employment grade) was inversely associated with risk of morbidity under a wide range of diseases, and with risk of mortality. That is to say that there was a social gradient observed, in which lower grades of civil servants had higher mortality and morbidity rates.

A follow up of the Whitehall II study showed that there was a social gradient present in health outcomes concerning sickness absence too.[17] For example, it was shown that men in the lowest grade of employment had a short term rate of absence 6.1 times higher than men in the highest grade of employment, and also 6.1 times higher for a long term sickness absence. The corresponding rates for women were 3.0 and 4.2 times higher, for short- and long-term rates of absences respectively.

1.2.3 Strategic Review of Health Inequalities in England (SRHIE) 2010 Report

One of the forefront reports of health inequalities in more recent times, was the Strategic Review of Health Inequalities in England post-2010 (SRHIE, 2010) report (or 'The Marmot Review').[18] Following the implications of the WHO CSDH report (2008),[12] the Secretary of State for Health requested that Professor Sir Michael Marmot assemble a team to research and devise effective strategies to reduce the specific health inequalities present in England.

Marmot and his colleagues showed that, whilst inequalities in mortality and morbidity were not as severe in England as they were globally, they still gave cause for concern. For instance, it was shown that those living in the least deprived neighbourhoods of England, were expected to live on average seven years more, than those from the most deprived neighbourhoods. Furthermore, those living in the least deprived neighbourhoods were also advantaged by an average of seventeen extra years of disability free life expectancy compared to those from the most deprived neighbourhoods. Thus, those from more geographically deprived areas in England were shown to experience shorter lives with more disability.

In terms of the impact on human years, it was shown by analysis reported in the SRHIE (2010)[18] that if the death rates in England were the same for all members of society as that of the most well off, then those who experienced a premature death would have benefited from between 1.3 and 2.5 million total extra years of life,[19] and also had 2.8 million extra total years of disability free life.[19]

Marmot et al also drew attention to the existence of a social gradient in health, namely that the association of social circumstances and health outcomes was not only observed when comparing the most against the least privileged in England, but that this association was a graded one. Where "put simply, the higher one's social position, the better one's health is likely to be".[18(p16)] As a result, Marmot et al called for social action to be universal, across all social strata, so as to reduce the health plight of all, rather than solely focus on the most vulnerable (with the caveat of ensuring that social actions are proportional to the level of disadvantage of each social group). The hope was that this would at the least reduce the steepness of the social gradient, given that eliminating it altogether is highly ambitious.

An important observation from the SRHIE (2010)[18] was that to reduce health inequalities, action is required against all the social determinants of health. The array of such social determinants of health is vast and complex, and hence this is a difficult problem to solve completely, involving inequalities in: early child development, education, occupation, neighbourhood, and housing conditions, to name but a few. More generally speaking, tackling this important issue of health inequalities, is about giving equal opportunities to all, allowing all members of society to equally receive the benefits society offers during their entire life course, from the moment of birth through to death.

It must also be noted that to have maximum effect, these actions cannot be implemented in isolation by the Department of Health and NHS alone, but must instead be a collaborative effort, involving the additional support of central and local Government, the third and private sectors, and local community groups.

Aside from impacting on quality and length of life, addressing social determinants of health inequalities has a reach beyond health-related benefits alone. Namely, there are important consequences for the economy. In the SRHIE (2010)[18] it was estimated that productivity losses due to health inequality were in the region of £31-33 billion per year,[19] with the loss of taxes and required additional welfare payments also costing a further £20-32 billion per year.[19]

It was also shown that over 75% of the English population did not have a disability free life expectancy that reached the age of 68 - the proposed English pension age in the coming years.[20] This signifies that if society expects to have a healthy population that works through to retirement age, then much work needs to be done to address the social determinants of health inequality alongside improving overall levels of health too.

Further consolidating this point, was a study by Parker et al (2020)[21] that investigated healthy working life expectancy (HWLE), defined as the mean number of years for a person to be expected to be healthy and in work post 50 years. The authors found that HWLE in England did not reach the State Pension Age (SPA): from age 50 there was shown to be an average of 9.4 years of HWLE – considerably less than the remaining years to the current and future proposals around SPA. Inequalities were also present by socioeconomic status, geographical region, and occupation. For example, HWLE was 1.5 times higher on average in the least deprived, compared to the most deprived quintile of England's population.[21]

One of the foremost policy recommendations of the SRHIE (2010),[18] which relates directly to these economic consequences of health inequality, concerns employment. Marmot et al[18] set as a specific policy objective, the need to improve access to and quality of good employment, across all social strata. Additionally, they emphasized the need to reduce long-term unemployment across all social strata. This is important because it is known that work is generally good for physical and mental health, and worklessness is associated with poorer health and well-being.[3] Hence improving inequalities around employment has potential implications for reducing health inequalities.

1.2.4 SRHIE 10 Year Follow-Up Report (2020)

A follow up of the SRHIE (2010)[18] was also conducted 10 years later, named 'Health Equity in England: The Marmot Review 10 Years On'.[22] This finds little positive change in tackling health inequalities. Marmot went on record to say that since the SRHIE (2010) "Britain has lost a decade. And it shows"[23(p1)].

Life expectancy at birth, which is a key marker of the health of a society and a measure that has been increasing since the start of the 20th century, has slowed in its growth dramatically from 2011. For example, from 1981 to 2010 there was a constant steady improvement in life expectancy (for men: a 1 year increase every 4 years, and for women: a 1 year increase every 5.5 years), which slowed significantly in 2011 through to 2018 (to just a 1 year increase every 15 years for men, and a 1 year increase every 28 years for women).[24] Life expectancy had even decreased in certain parts of England, and generally health inequalities had widened.

Although the SRHIE (2010) and its resultant policy recommendations were originally welcomed by the UK Government, the austerity of the subsequent decade, which saw public expenditure of national income reduced from 42% in 2009-2010 to 35% in 2018-2019, affected almost all of the policy areas recommended by Marmot et al negatively.[22]

For example, funding for education had been reduced, child poverty had increased, and more people had inadequate money to live a healthy lifestyle and needed to turn to food banks for support. Marmot et al hypothesized that this deterioration of the quality of society "is likely to have had an adverse effect on health and health inequalities".[23(p1)] Thus the need to act on reducing the harmful effects of social determinants on health inequalities remains prominent today.

A study by Hiam et al (2020)[25] was also in agreement that health improvements had slowed in the UK since 2010, particularly with regard to life expectancy and infant mortality. The study authors proposed that greater investments be made towards researching determinants of health, and that policies are implemented to improve the health and well-being of all.[25]

1.2.5 Health Inequalities and the COVID-19 Pandemic

The impact of more contemporary events relating to the COVID-19 pandemic on health inequalities have also been explored by Michael Marmot's team via the 'Build Back Fairer: The COVID-19 Marmot Review',[26] published in December 2020, using the 'Health Equity in England: The Marmot Review 10 Years On'[22] as a base.

In this COVID-19 Marmot Review, the notion of building society back to be fairer post pandemic, rather than solely better, was advocated.[26] Marmot et al suggested that a more just distribution of health and wellbeing be the key focal point of building a fairer society, and that the Government should increase its prioritization of action against the social determinants of health. The pandemic played a pivotal role in compelling the Government to act against certain health inequalities in the short term – including those around poorer school children more likely to be malnourished, and rough sleepers dying prematurely. However, longer term policies to reduce health inequalities based on equity were recommended, alongside more investment in public health.

Marmot et al showed that the inequalities prior to the pandemic in relation to social and economic conditions, led to higher and more disproportional mortality rates due to COVID-19. For example, higher mortality rates due to COVID-19 were observed from people in the Black, Asian and Minority Ethnic (BAME) group than those of white ethnicity, and it was suggested that contributing factors for this were because BAME community members were more likely to live in disadvantaged neighbourhoods, work in higher risk occupations, and live in overcrowded conditions – a result of longstanding inequalities.

1.2.6 Worklessness Drives Health Inequalities

Thus far, health inequalities and a broad range of their corresponding social determinants have been discussed. For the remainder of this thesis, the focus is now narrowed down to just one of the known drivers of widening health inequalities – long-term sickness absence, through the ability to work or worklessness (a state of unemployment or economic inactivity).[7]

Being without employment is deleterious to an individual's ability to access money, power and resources; factors which are key social determinants of health.[27] It has also been stated that "work is the most important determinant of population health and health inequalities in advanced market democracies".[28(pIX)]

Patterns of employment and work absence over time both reflect and reinforce the social gradient, and demonstrate the inequalities of access to labour market opportunities. People who are in lower socioeconomic positions are at a greater risk of unemployment,[29] and being unemployed is associated with a greater rate of long-term illness,[30] as well as mental illness.[31] In contrast, being employed provides psychological benefits such as social inclusion and identity, a sense of purpose, and regular activity/routine to one's life.[4] However, a caveat is that these beneficial health effects of work are dependent on the nature of the job itself. For example, jobs that involve demands greater than employees' capabilities can have the reverse effect and contribute towards poorer physical and mental health.[3]

Ill health itself has also been shown to contribute to worklessness, hence the problem is cyclical.[32] For the majority of the healthy working population the spiral towards worklessness tends to start with the onset of ill health, if this progresses to a point in

which sick pay is required through the state, a fit note will usually need to be obtained from a healthcare professional (HCP), for example, a General Practitioner (GP). Fit notes are explored in the following section.

1.3 Fit Notes

1.3.1 Definition and Purpose of Fit Notes

A Statement of Fitness for Work, more simply known as a 'fit note', is a written statement from a HCP that records the medical advice a patient has received regarding their fitness to work.[33] The premise behind a fit note is that the right kind of work is generally good for a patient's physical and mental well-being,[3] and that it is not necessary to be completely fit to work in many instances.[34] Hence fit notes are administered with the intent of helping patients RTW as soon as they can and aiding their recovery.[35,36]

In the UK, a fit note is issued free of charge, after being off work for more than seven consecutive days due to a medical problem. If a patient is off work for seven days or less, no medical evidence is required to confirm this absence to employers. If a patient's fitness for work is affected by a non-medical problem, such as problems at home, a fit note cannot be issued (although other forms of support can be administered).[37]

A fit note can be issued by the following HCPs: doctors, nurses, pharmacists, physiotherapists, or occupational therapists. Although, latest fit note issuance data from NHS Digital shows that the majority (91.6%) of fit notes are issued by doctors, with only 6.5% issued by nurses, and < 2% by the remaining HCPs.[38]

1.3.2 Change in UK Sickness Certification System (2010)

Prior to the fit note being introduced, another sickness certification system was used in the UK, but in 2008, a UK Government report titled 'Working for a Healthier Tomorrow' by Dame Carol Black proposed that this old sickness certification system was in need of a change.[39] Thus, on 6th April 2010, the old medical statements (Med 3 and Med 5) which had been in use since 1948, were replaced with an amalgamated form of the two - the fit note.[40]

One of the fundamental issues with the old system, as identified by Black et al (2008),[39] was that it had the underlying assumption that being ill and being in work were mutually exclusive events, and the misconception that working impedes recovery. The old sick note only required the HCP to briefly state their patient's health condition and the quantity of time they could expect to be absent from work as a result. The emphasis was on what the person could not do, rather than what they could (for example, no regard was given to amending workplace duties).

The shift in attitude in workplaces, from an age where employers had more rigid expectations around employees being able to perform certain duties, to the modern age where more flexible working is acceptable, also played a role in facilitating this change to occur.

The key difference of the fit note from the old sickness certification system, is that the fit note additionally allows the HCP to recommend that the patient 'may be fit for work' (and hence extends the old sick note binary system of 'fit to work' or 'not fit to work'). In the fit note, if the patient 'may be fit for work', the HCP can recommend that they consider (along with their employer): 'a phased return to work', 'amended duties',

‘altered hours’, or ‘workplace adaptations’.[36] Furthermore, documented within the fit note are the details of the functional effects of the patient’s condition (again for the benefit of the patient and the employer to consider together), this may include limitations relating to, for example, stamina, mobility, and cognitive abilities.[35]

In general, it is expected that the HCP will have a discussion with the patient about what it is possible for them to do at work alongside their health condition, rather than declaring that the patient is not fit for work by default. The guidance given by the HCP is not occupation specific, but rather about the general fitness for work of the patient. Hence the next vital step after receiving a fit note, is that patients and their employers should work together to consider possible mechanisms to aid their RTW.[33]

The discussion of sickness absence in this Section primarily applies to only employed individuals. Self-employed individuals, who work for themselves instead of an employer, can still be issued a fit note, but are not eligible to claim sickness absence remuneration through the SSP system. However, self-employed individuals may be entitled to claim other support, such as Employment and Support Allowance (ESA) or Universal Credit.[5]

1.3.3 Longer Term Work Absences Impede RTW

The old sickness certification absence system, due to its lack of emphasis on encouraging presenteeism (under specific conditions where it is deemed a medically beneficial option for the patient’s recovery), was thought to contribute to the issue of long term sickness.[41] In contrast, the introduction of the fit note was thought to reduce the risk of patients moving into long term work absence.

Although research shows that the majority of people do RTW within a short time frame following a medical problem, around 10% are expected to go on to experience longer term absences of > 12 months.[9] The lengthier the work absence, the harder it is for the individual to RTW, and the more adverse health and social problems they experience.[3] Indeed, a report by the ONS showed that during 2021 and 2022, only 16% of individuals who were on a long-term sickness absence had a RTW.[1]

One reason for the difficulty in returning to work after a lengthy absence is due to loss of confidence, something that can intensify with lengthier durations of work absence. Anxiety is the most often mentioned impediment to a RTW after an episode of ill health.[42] Whereas the actual medical problem itself, or the employee's self-perceived ability to manage their illness while at work, is often not cited as an impediment to RTW.[42]

To prevent loss of confidence and onset of anxiety, one of the most important factors in facilitating a timely RTW, is regular contact and maintaining involvement with the person's workplace during their illness.[43] Another important factor to promote a timely RTW, even whilst the employee may not be fully fit, is adjustments to the workplace (conditions, duties, hours etc).[44]

1.3.4 Defining Duration of Sickness Absence

HCPs prescribe fit notes for a period of time in accordance with their best judgement, this could be: days, weeks, or even months. The initial time on the first fit note can be extended by issuing another fit note, with its own corresponding time frame. Whilst there remains debate in defining long- and short-term sickness absence[45] (which in turn

makes comparisons between studies and countries difficult)[46], guidance from NICE[47] does provides the following definition:

- Short term sickness absence: lasting up to four weeks
- Recurring short term sickness absence: number of episodes of absence from work, with each lasting less than four weeks
- Long term sickness absence: four weeks or more

1.3.5 Performance of the Fit Note

The initial introduction of the fit note in 2010 was generally favoured by GPs,[48] and a Department for Work and Pensions (DWP) Report, titled 'General Practitioners' attitudes towards patients' health and work, 2010–12',[49] showed that 60.5% of 1405 GPs thought that the introduction of the fit note had increased their likelihood of recommending patients RTW as a means of recovery, in a 2012 survey.

However, in practice, there does not seem to be any tangible evidence suggesting that fit notes have thus far reduced sickness absence. In a 2018 systematic review that evaluated the effectiveness of the fit note in the UK, it was shown that in the largest included study, only 6.5% of fit notes administered made use of the novel 'maybe fit for work' option.[50] More recent data from NHS Digital, similarly showed that 6.4% of all fit notes issued from January to September 2023 were issued as 'maybe fit for work'.[38]

The systematic review study authors mentioned that patients may feel stigma in disclosing their health condition on a fit note, as they are aware that their employer will be able to access this information, which could then be contributory in a reluctance from the patients' side to engage with the 'maybe fit for work option'.[50] In general, the study

authors recommended that further research was necessary to better understand the success of the fit note.

Furthermore, a 2017 DWP report, 'Improving Lives: The Future of Work, Health and Disability',[51] stated in their internal review of the fit note that "the fit note remains an important tool, but it is not always used effectively across the system to support people staying in or returning to work",[51(p43)] and that there are still too many fit notes that "say 'not fit for work' when people 'may be fit for work' as long as appropriate workplace adjustments are made".[51(p43)] In response to this, they implemented actions such as extending fit note certification authorization to other HCPs (aside from only GPs), and integrated specialized training on the fit note in the GP training pathway.

Some GPs themselves remained sceptical too and considered employers to be the primary hindrance to an early RTW.[36] Unlike in Nordic countries, in the UK, employers are not legally obligated to comply with GP fit note recommendations, hence this can lead to tensions between GPs and employers.[52] However, when the GP recommended RTW adjustments are clear, there is evidence suggesting that employers are more receptive to this request.[50]

Although the performance of the fit note is unclear, the main contributors to sickness absence are clear – musculoskeletal (MSK) and mental health (MH) conditions. These are discussed in the next section.

1.3.6 Musculoskeletal and Mental Health Conditions

Poor health, and in particular chronic conditions, significantly affect one's ability to work. Chief among these are MSK and MH conditions. For example, the 'Mental health at work: The business costs ten years on' report estimated that in the 2016/17 financial

year £10.6 billion was spent on MH related sickness absence in the UK.[53] Additionally, the 'Work related musculoskeletal disorder statistics (WRMSDs) in Great Britain, 2020' report showed that MSK disorders caused by work accounted for 27% of total working days lost due to ill health in 2019/20, and that 8.9 million working days were lost due to this condition in this time period (these figures will be further increased when extending to consider all MSK condition sickness absences, regardless of cause).[54]

Over the past few decades, patterns of UK sickness absence have changed markedly.[55] In particular, while MSK and MH conditions continue to be the predominant reasons for long-term sickness absence, there has been a trend in mental illness overtaking MSK conditions as the primary cause.[56]

For example, a study aiming to report sickness certification rates in a UK population based in North Staffordshire, found that in 2005, under the old system of sick notes, sickness certification rates were indeed greatest for patients with MSK and MH conditions.[57] However, the rate of sickness certification was slightly higher for those with MH conditions (27.78 per 1000 person years), than those with MSK conditions (22.84 per 1000 person years).

Additionally, another study, also investigating changes in sickness certification in a UK population of North Staffordshire, showed that the rate of sickness certification due to back pain fell from 376.8 per 1000 back pain consultations in 2000, to 246.5 per 1000 back pain consultations in 2010.[55] Although exact reasons for this decline in back pain sickness certification are unclear, the study authors suggested it could be due to shifts in advice around managing back pain, with guidelines during the study time period (2000-2010) encouraging individuals to maintain an active lifestyle and RTW, as far as possible, despite their back pain.[55]

More recent data shows that MH conditions, followed by MSK conditions, are still the predominant reasons for sickness absence in the UK. For example, the September 2023 fit note issuance data from NHS Digital relating to GP Practices in England, showed that MH conditions comprise the highest percentage of total fit notes issued (out of the fit notes issued in England that contained a medical classification under the widely used 10th revision of the International Statistical Classification of Diseases and Related Health Problems (ICD-10)), with 37.5% of fit notes issued due to 'mental and behavioural disorders', and the second most predominant reason was a MSK condition, with 17.9% of fit notes issued due to 'Diseases of the musculoskeletal system and connective tissue'.[38]

The rise of MH problems driving sickness absence was also observed in a 2016 study by Gabbay et al.[58] Here sickness absence data from 68 general practices in the UK was analysed, using 25,078 fit notes and 13,694 corresponding sickness episodes. Common mental disorder (CMD) was the reason for 29% of all sickness episodes. Stress was the most prevalent type of mental disorder according to the total number of fit notes analysed, and depression (with or without anxiety) had the greatest impact on the total duration of the sickness absences caused by CMDs (47% of the total fit note weeks due to CMD). Gabbay et al classed 'long-term' sickness absence as 12 weeks or more, and 16% of CMD episodes fell into this category, with older people being the most likely to have a long term CMD sickness absence.

In agreement with the low use of the 'may be fit' option discussed in the previous Section, it was also noted by Gabbay et al (2016)[58] that only 7% of the CMD fit notes had this option recorded, and the patients least likely to receive this type of fit note for their CMD were those living in more deprived neighbourhoods. Gabbay et al

recommended that to facilitate a timelier RTW, GPs should consider using the 'may be fit for work' option in a more effective manner for CMD fit notes and consider referring patients to the recently established Fit for Work service.

In addition to this, a study based on a Norwegian population, advocated that with regard to CMD sickness absences, a better insight from GPs could be more effective in encouraging RTW, as lower risks of sickness certification and prolonged sick leave due to CMD were observed for patients with specialist GPs.[59]

1.4 Tackling the Issue of Sickness Absence

Thus far the significance of the problem that sickness absence poses in the UK has been discussed.

There are several stakeholders involved in the process of tackling RTW, including:

- The Individual
- The Employer
- The Government
- Health Services (e.g., NHS)

Whilst each stakeholder plays a role in tackling this problem, the management of the RTW is generally considered to be a shared responsibility between these stakeholders.

The role of MSK and MH conditions in perpetuating the work absence cycle have also been noted in the previous Section, thus the question remains how to tackle this issue? One of the routes is through access to fit note data, as will be discussed in this Section.

1.4.1 Fit Note Data

Fit notes are recorded in primary care as electronic health records (EHR). There are several large anonymised national databases of EHRs available for research, such as the Clinical Practice Research Datalink (CPRD). The CPRD is a longitudinal national database of primary care EHR that contains data from over 60 million patients across the UK.[60]

Increasing access to primary care EHRs provides a unique opportunity to examine work absence, by assessing patterns in the issuance of fit notes. During an initial HCP consultation for sickness absence, it is challenging to determine which patients are at the highest risk of sustained long-term work absence; determining trajectories of work absence and the patient characteristics associated with them may assist with this.

First, the concept of longitudinal trajectories is introduced.

1.4.2 Defining Longitudinal Trajectories

A trajectory describes the evolution of a repeated measure over time (for example, the course of work absence). This is achieved by identifying subgroups of individuals with similar patterns in a set of longitudinal heterogeneous data. Trajectory modelling is thus focused on relationships among individuals, and is designed to assign individuals to subgroups in accordance with their personal response patterns. Classification into subgroups is performed such that individuals share more similarities within their subgroup, than outside of the subgroup.[61]

Studies often tend to average out the effects of health outcomes and their course, either across the whole study sample or observed subgroups that are specified a priori. However, patterns of behaviour and clinical symptoms for example, are often shared by

individuals from subgroups that are unknown or not pre-identified. Therefore, by using averaged estimates to describe populations of individuals, and ignoring the complex intra- and inter-individual variability of the real-life clinical context, much useful information is lost and conclusions can be misguided.[62]

Methodology incorporating trajectory modelling offers a solution to this problem and allows individuals to be designated to homogeneous subgroups (i.e., distinct trajectories), in accordance with similarities on given outcomes. This could be, for example, different subgroups related to behaviour around work absence.

1.4.3 Application of Trajectory Analysis

Trajectories can be applied to work absence, for example, to investigate how patients experiencing a baseline work absence subsequently behave over time. There might exist one subgroup that has a high probability of a quick RTW and without a subsequent relapse into work absence, another subgroup that experiences intermittent periods of RTW and work absence, and a final subgroup that has a high probability of no RTW at all. Many more complex combinations of work absence patterns also exist, which trajectory models can also handle. Furthermore, once trajectories of work absence are established, it is possible to explore whether there are identifiable and modifiable risk factors, that are associated with following a more severe course of work absence over time.

Trajectories of the course of health conditions are common in today's literature. For example, studies exploring trajectories of low-back pain were set into motion by an innovative 2006 study[63] which used a statistical approach known as latent class analysis

(LCA; different methods to derive trajectories are discussed in Chapter 5 of this thesis) to identify four different courses of low-back pain over time:

- Class 1 ("persistent mild") whereby patients had stable, low levels of pain
- Class 2 ("recovering") where there was initial mild pain, which progressed rapidly to no pain
- Class 3 ("severe chronic") whereby patients had continuously high pain
- Class 4 ("fluctuating") where pain altered between mild and high levels

A review into low back pain trajectories research in 2016 identified ten such studies, which despite showing some differences between studies also presented several common trajectories of low back pain, and supported the idea that modelling trajectories of a health condition can be useful to improve understanding of the condition and its corresponding clinical management, which is appealing to clinicians and can also be helpful as a communication tool for patients.[64]

Another study, of low-back related leg pain, used LCA methods to identify four pain trajectories: improving mild pain; persistent moderate pain; persistent severe pain; and improving severe pain.[65] Statistically significant differences of trajectory membership depending on baseline characteristics were shown by the study authors.[65] Furthermore, Nicholls et al (2014) found five pain trajectories in their study for individuals with, or at high risk of knee osteoarthritis.[66]

Similarly, for MH conditions there exists a breadth of trajectory literature, such as a 2019 study on the trajectories of MH amongst women in Australia as they age,[67] and a 2021 study of MH trajectories in an English population during the COVID-19

pandemic.[68] In general, such trajectory research has uncovered novel pathways into MH and MSK conditions.

1.4.4 Importance of Trajectories of Work Absence

However, trajectories of work absence are scarcely studied. Much of the work absence literature has been focused on identifying risk factors for experiencing a sickness absence; few studies consider individuals already on a sickness absence at baseline (which is important as this is the context in which a HCP receiving a patient at first consultation for sickness absence must operate).[69]

Recent studies are starting to assess trajectories of work absences for patients absent from work at baseline due to: MSK conditions (e.g. in: a Canadian population;[70] a Dutch population and only considering arm, neck and/or shoulder complaints;[71] and a Swedish population and only considering osteoarthritis),[69] or MH conditions (e.g. in a: Swedish population and only considering depression;[72] and a Dutch population).[73]

However, this remains a niche area, and to the best of our knowledge, our study is the first to explore trajectories of work absence due to a MSK or MH condition in an English population, as well as the profiles of patients associated with such trajectories. This is important because a more complete understanding of the intricate courses of work absence over time is hypothesised to be useful in tackling the issue of work absence. Hence, use of trajectories in this context may help HCPs to better understand, at initial consultation, which patients are more likely to undergo a detrimental course of work absence over time, and thus to act on this information by providing more timely and targeted support for such patients.

1.5 Conclusion

To conclude, in this Chapter, first an overview of long-term sickness absence was provided and the need to reduce it. Then, the social determinants of health inequalities were discussed through various examples, as well as the extensive adverse effects of health inequalities for the individual, the economy, and wider society.

Furthermore, consideration was given to the issue of worklessness driving health inequalities. On the contrary, it was observed that working is generally beneficial for health (albeit depending on the type of employment), and that working whilst not fully fit can promote recovery too.

The path into worklessness often starting with an initial sickness absence was discussed, and how this can lead to a longer-term absence, and that these initial sickness absence consultations are recorded as fit notes. Then, MSK and MH conditions as the primary causes for sickness absence were explored.

Finally, the availability of fit note data, which allows investigation of trajectories of work absence was discussed, and the benefits that this can provide in identifying which individuals are at greater risk of a long-term work absence, so that they could receive earlier and more targeted intervention, which may increase their likelihood of recovery.

In the next Chapter, the research questions that have arisen from this background Chapter are first condensed into specific aims. Then, an overview of the remaining Chapters of this thesis is provided, in which these thesis aims are answered.

Chapter 2. Thesis Aims and Structure

In this Chapter, the overall aims and the structure of this PhD are detailed.

2.1 Aims

This PhD has two main aims:

1) To derive common longitudinal trajectories of work absence as measured by receipt of fit notes, for a population consulting their HCP with a MSK or MH condition
2) To identify health and sociodemographic characteristics associated with these trajectories

To address these two aims, a systematic review and three original research studies are performed in this thesis, as summarised in the next Section.

2.2 Thesis Structure

Firstly, in the next Chapter, prior to performing any original research in this thesis, a systematic review is conducted to evaluate what is already known in the existing literature concerning trajectories of work absence due to a MSK or MH condition.

Then, in Chapter 4, different datasets suitable for performing the original research in this thesis are described and contrasted, and a final dataset is chosen. Finally, a last step before performing any analyses is to describe and compare appropriate statistical methods for modelling longitudinal data in this thesis – this is outlined in Chapter 5, with a detailed review of different trajectory derivation methods and their application.

In Chapter 6, the first of three original research studies of this thesis is conducted to understand trends concerning incidence rates of fit notes issued due to a MSK or MH condition. The specific research questions for this study are:

1. What is the annual rate of work absence in individuals consulting their HCP with MSK or MH conditions?

    1.1 Do these rates of work absence vary by the following sociodemographic characteristics: age, sex, and geographic region?

Then, the main analyses of this thesis is performed in Chapter 7 to address the first thesis aim (from Section 2.1). In this study, the baseline population is taken as a subset from those with incident fit notes due to a MSK or MH condition as identified in Chapter 6, and the individuals are then followed up for trajectory derivation analysis. The following research questions answered:

2. Is it possible to identify longitudinal trajectories of work absence in individuals presenting to their general HCP with a MSK or MH condition?

    2.1 What are the most appropriate time intervals for determining trajectories of work absence?

    2.2 What is the most appropriate method for determining trajectories of work absence?

    2.3 Are the derived trajectories of work absence dependent on reason for index fit note?

The final study performed in this thesis is then presented in Chapter 8. Here, the optimal trajectories of absence due to a MSK or MH condition, as identified in Chapter 7, are

assessed for association with an array of sociodemographic and health characteristics, as well as types of treatment received and comorbidity. This study answers the second thesis aim (from Section 2.1), and addresses the following questions:

3.1 What are the typical profiles of individuals within each of the identified work absence trajectories?

3.2 Is it possible to identify health and sociodemographic characteristics associated with future persistent or recurrent work absence?

3.3 Do the typical profiles, and any observed associations of characteristics with work absence trajectories differ by reason for index fit note?

Finally, in Chapter 9 a discussion of the key overall findings resulting from this thesis are presented, as well as areas for further research.

Chapter 3. Systematic Review of Longitudinal Trajectories of Work Absence in Individuals With Musculoskeletal and/or Mental Health Conditions

3.1 Introduction

Chapter 1 illustrated that work absence can have significant consequences for an individual, their employer, and wider society, and that MSK and MH conditions were the principal reasons for such absences. Furthermore, it was hypothesised that derivation of common trajectories of work absence, which encapsulate the changing course of work absence patterns over time, were important to aid HCPs in better identifying potential work absence patterns for their patients at initial consultation for sickness absence, which could thus help HCPs to provide earlier and more targeted inventions to support their patients.

In this Chapter, a thorough search and critique of the existing literature is performed via a systematic review to more comprehensively understand the current knowledge base and literature gaps in relation to trajectories of work absence.

Systematic reviews allow the literature for a particular research problem to be searched, critically appraised, and synthesized in an effective manner.[74] By transparent reporting of the methodology used, the results of a systematic review are usually expected to be closely reproducible. Systematic reviews are increasingly used and regarded as a gold standard in evidence based medicine and are especially useful due to the objective manner in which they allow large quantities of research to be summarised.[74]

The standard steps when performing a systematic review are the following:[74]

1. Clearly defining the research question of interest.
2. Formulating a strategy to search the literature for potentially useful studies in scientific resources, such as electronic databases.
3. Screening the retrieved literature to retain only the most relevant studies to the pre-determined research question (this is based on inclusion and exclusion criteria that is decided a priori).
4. Appraising the quality of studies that have been included post screening (relevant appraisal tools/checklists can be used to aid this process).
5. Extraction of the most important and relevant data from included studies using a pre-defined data extraction form.
6. Synthesizing and discussing the findings from included studies using a meta-analysis (for quantitative data), or where this is not possible, a narrative review.

3.2 Objectives

This systematic review had the following objectives:

1) To determine published longitudinal trajectories of work absence in individuals with baseline work absence due to a MSK and/or MH condition.

2) To determine presence of any reported trajectory-covariate associations.

3) To critically appraise the analytical methods used for modelling longitudinal trajectories of work absence in published studies of individuals with baseline work absence due to a MSK and/or MH condition.

3.3 Methods

A protocol for this systematic review was developed a priori, and with reference to the Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P) reporting standards.[75] AL developed the protocol, and amended following advice and feedback from the Systematic Review Team at Keele University.

3.3.1 Search Strategy

After having defined clear research objectives for this systematic review, the next step in the process was to construct a well-defined search strategy, first by selecting appropriate research databases to search. Table 1 shows details of the databases that were searched for relevant literature in this systematic review, and under which platform they were searched. These databases were chosen to best represent the aims of this review and were agreed upon with the expert advice received from the Systematic Review Team at Keele University. These databases cover fields relating to: ageing research, medicine, biomedicine, social sciences, behavioural research, psychology, nursing, and more.

An initial search was conducted from inception of each database up to 8th April 2021 (an updated search was also performed later and is described in Section 3.3.8).

Table 1. Description of Databases used in Systematic Review Search Strategy

Database	**Searched Using:**	**Description**
AgeLine	EBSCO	Focuses on issues relating to ageing in people aged over 50.
APA PsycInfo	EBSCO	Relevant for studies relating to behavioural and social sciences, from a psychology perspective.
APA PsycArticles	EBSCO	Contains full text articles published by the American Psychological Association, and affiliated journals. Encompasses full range of psychology, and many allied fields such as medicine, nursing, and public health.
CINAHL Plus with Full Text	EBSCO	A large collection of open access journals relating to nursing and allied health.
Allied and Complementary Medicine (AMED)	OVID	Covers journals in: complimentary medicine, palliative care, and other allied health professions (physiotherapy, occupational therapy, podiatry, rehabilitation, and speech & language therapy).
EMBASE	OVID	Contains international biomedical and pharmacological literature.
MEDLINE	OVID	Designed to cover biomedicine and health, encompassing: life sciences, chemical sciences, behavioural sciences and bioengineering.
Science Citation Index Expanded	Web of Science	Covers over 178 scientific disciplines, including science and medicine.

A search strategy was developed with combinations of search terms constructed based around the four concepts relevant to this systematic review: work absence, MSK conditions, MH conditions, and longitudinal cohort studies that include trajectory analysis.

The search strategy design was largely similar across the databases, with slight modifications applied to account for nomenclature differences across the databases when conducting free-text searches, as well as differences in subject headings relating to the medical thesauruses relevant to each database (e.g., Medical Subject Headings (MeSH) is the thesaurus used for OVID and has its own specific set of subject headings).

To identify as much applicable literature as possible, for each concept, each search strategy involved first attempting to identify as many relevant subject headings as possible from the medical thesaurus attached to the database, and then searching for any further terms related to each concept through free-text searches of titles and abstracts (the search strategy for the Science Citation Index Expanded database using Web of Science was the only exception, as there are no thesauruses attached to this database, hence only free text searching was possible).

Decisions were also made around which subject headings to 'explode' – this allows the search to include all child subject headings (i.e., subject headings that fall under the same branch as the main subject heading) – by reviewing the list of child subject headings relating to each identified main subject heading, and deciding on the relevance of each to this thesis; this process, as with the whole search strategy generation process in general, was guided by the advice received from the Systematic Review Team.

Database searches were conducted by a single reviewer (AL).

An example of the search strategy in the MEDLINE database, run using the OVID interface, is provided in Table 2 (and a more complete version that includes the number of studies identified at each stage in Appendix A).

Table 2. Systematic Review Search Strategy using MEDLINE (OVID)

	Concept	Search Term
1	Work Absence	ABSENTEEISM/
2	Work Absence	(absenteeism).ti,ab
3	Work Absence	exp Rehabilitation, Vocational/
4	Work Absence	((job OR work OR occupation* OR vocat*) ADJ2 (absen* OR rehab* OR adj* OR participation OR incapacity OR leave OR return)).ti,ab
5	Work Absence	Sick Leave/
6	Work Absence	(sick* ADJ2 (leave OR absen* OR day OR note OR cert* OR pay* OR paid)).ti,ab
7	Work Absence	(fit* ADJ2 note).ti,ab
8	Work Absence	Return to Work/
9	Work Absence	("return to work" OR rtw OR (work ADJ2 resumption) OR (work ADJ2 re-entry) OR (work ADJ2 return)).ti,ab
10	Work Absence	((long-term OR "long term" OR longterm) ADJ2 (sick* OR abs*)).ti,ab
11	Work Absence	((job OR work OR occupation*) ADJ2 (ability OR able OR disab* OR capacity)).ti,ab
12	Work Absence	(incapacity).ti,ab
13	Work Absence	1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12
14	Musculoskeletal Health Conditions	exp Musculoskeletal Diseases/
15	Musculoskeletal Health Conditions	(musculoskeletal ADJ2 (disease* OR pain OR injur* OR disorder OR disorders OR condition*)).ti,ab
16	Musculoskeletal Health Conditions	exp Pain/
17	Musculoskeletal Health Conditions	(pain ADJ2 (back OR lumbar OR hand OR knee OR "joint chronic" OR persistent OR "long term" OR long-term OR longterm OR widespread)).ti,ab

18	Musculoskeletal Health Conditions	(pain ADJ2 low* back).ti,ab
19	Musculoskeletal Health Conditions	(arthr* OR osteoarthr*).ti,ab
20	Musculoskeletal Health Conditions	14 or 15 or 16 or 17 or 18 or 19
21	Mental Health Conditions	Mental Health/
22	Mental Health Conditions	exp Mental Disorders/
23	Mental Health Conditions	(mental ADJ2 (health OR disorder* OR illness*)).ti,ab
24	Mental Health Conditions	exp Anxiety/
25	Mental Health Conditions	(anxiety).ti,ab
26	Mental Health Conditions	Depression/
27	Mental Health Conditions	(depression).ti,ab
28	Mental Health Conditions	(psychiatric ADJ2 illness*).ti,ab
29	Mental Health Conditions	21 or 22 or 23 or 24 or 25 or 26 or 27 or 28
30	Longitudinal Cohort Studies (with Trajectory Analysis)	cohort studies/ or follow-up studies/ or longitudinal studies/ or prospective studies/
31	Longitudinal Cohort Studies (with Trajectory Analysis)	(longitudinal or prospective or cohort).ti,ab
32	Longitudinal Cohort Studies (with Trajectory Analysis)	Observational Study/
33	Longitudinal Cohort Studies (with Trajectory Analysis)	(observational).ti,ab

34	Longitudinal Cohort Studies (with Trajectory Analysis)	(trajector* or pattern*).ti,ab
35	Longitudinal Cohort Studies (with Trajectory Analysis)	Latent Class Analysis/
36	Longitudinal Cohort Studies (with Trajectory Analysis)	(latent ADJ2 (class or transition)).ti,ab
37	Longitudinal Cohort Studies (with Trajectory Analysis)	30 or 31 or 32 or 33 or 34 or 35 or 36
38	Musculoskeletal or Mental Health Conditions	20 or 29
39	Longitudinal Cohort Studies (with Trajectory Analysis) of Work Absence, relating to Musculoskeletal or Mental Health Conditions	13 and 37 and 38

1. '.ti,ab' is used to search only in titles and abstracts
2. Subject Headings are searched for wherever '/' is used, this is in accordance with the medical thesaurus: Medical Subject Headings (MeSH)
3. Medical Subject Headings are exploded to include all child Subject Headings that fall in the same branch as the designated Subject Heading by use of the term 'exp'
4. An asterisk ('*') represents truncation. For example, occupation* searches for the terms: occupation, occupations, occupational etc.
4. Using quotation marks only allows the exact term to be searched (as specified in full in the text string)
5. 'ADJ2' allows up to two words to appear between the designated search terms (they can appear in any order)
6. The Boolean operator 'or' is used to combine the search using mathematical logic

Similar searches to the above were conducted across the remaining databases. Then, all the records identified from these searches were extracted and imported into reference management software, ready for the next stage in the process, de-duplication.

3.3.2 De-Duplication Prior to Study Screening

Mendeley Desktop software was used to de-duplicate results.

First, the automated Mendeley Desktop 'Check for Duplicates' tool was used to de-duplicate within each database separately. This tool searches for duplicates by identifying studies that match against any of the following criteria:

1) Study Title + Year of Publication,

2) Study Title + Abstract,

3) Study Title + Page Numbers,

4) Study Title + Author names,

5) Study Title + Digital Object Identifier (DOI).

Next, the remaining results were arranged alphabetically by study title, and de-duplicated by manual screening (within Mendeley Desktop). Differences in a word being capitalised and non-capitalised were the prime causes for most of the manually identified duplicates.

Finally, the remaining set of studies were once more checked with Mendeley's automated tool for further final de-duplication (now checking all remaining studies and not restricting to each separate database). Then, the unique (de-duplicated) references were imported into Rayyan screening software.[76]

De-duplication was conducted by a single reviewer (AL).

Prior to screening, several test searches were conducted in Rayyan software to ensure that previously identified relevant studies that were expected to appear in the list of studies to screen,[55,71,72,77,78] were present (i.e. this was a crude validation check of the search strategy itself).

3.3.3 Inclusion and Exclusion Criteria

In a systematic review, pre-stating the eligibility criteria that determines inclusion/exclusion of relevant studies is important. Pre-specification of this criteria is remarked upon as a notable differentiator between systematic and narrative reviews.[79] A useful aid in developing inclusion and exclusion criteria is to consider the PICO format (respectively: participants, intervention(s), comparator(s), and outcomes) where applicable, to isolate the different elements of the research question of interest, and also to consider the types of studies (e.g. randomised controlled trials) that will be relevant to the search.[79]

Having clearly defined inclusion/exclusion criteria a priori is imperative in guiding the screening process to allow only truly relevant studies to be selected, and to remove as much ambiguity in this process as possible, both for the original researchers conducting the systematic review, and for other researchers who may wish to validate or duplicate this search.

In this systematic review, the inclusion and exclusion criteria applied are as detailed in Table 3. These criteria were selected to represent the objectives of this systematic review as closely as possible and were applied to the screening of titles and abstracts (as a combined step), as well as full texts.

Table 3. Inclusion and Exclusion Criteria used in Systematic Review Screening

Inclusion	Exclusion
- Working age adult individuals or adults that are economically active, and absent from work at baseline due to a MSK and/or MH condition - Only longitudinal cohort studies (however systematic reviews of these are included if identified) - Studies which report a trajectory or pattern of work absence - Studies set in: primary care, community care, occupational healthcare, rehabilitation (where RTW interventions are delivered), or workplace settings (e.g. in the US, where Health Maintenance Organisations (HMOs) provide healthcare to large workplaces via the insurance system)	- Individuals who have experienced an accident outside of work that caused work absence (e.g., a motor vehicle accident that caused whiplash) - Individuals undergoing or who are stated to have undergone any type of surgery - Studies that only aim to identify factors/predictors of RTW, or only mediators/moderators of a work absence association (i.e., there is no trajectory of work absence present) - Participant populations of combinations of conditions at baseline, where at least one of the co-morbid conditions is not a MSK and/or MH condition (e.g., exclude if population is cancer individuals with rheumatoid arthritis) - Individuals who are receiving any form of intervention (that is in addition to normal care) - Studies with no full text available (at full text screening stage only, and exclusion only applies after exhausting all possible options to obtain full text) - Conference proceedings - Non-English language publications (where appropriate translation resource is not available) - Any duplicate studies that were not previously detected in the pre-screening de-duplication process

Abbreviations: MSK = Musculoskeletal; MH = Mental Health.

3.3.4 Screening for Relevant Studies

Screening (using the inclusion/exclusion criteria from Table 3) was done in two stages: first with titles and abstracts screened as a combined step (using Rayyan software[76]), and then as a second step, full texts.

AL performed the whole screening process, and to affirm the integrity of decisions made, a second reviewer (GWJ) also independently screened a random sample of 10% of the total studies at title/abstract stage, as well as all the studies at full text stage. Any discrepancies between the two reviewers were resolved through consensus meetings as required, with a third person (KJ) available to arbitrate if necessary.

Studies with no abstracts available, were carried forward to the full text screening stage for completeness, after first exhausting all best possible efforts to find the abstracts using online searches (by both AL and GWJ).

If full texts of studies included after title/abstract screening could not be found after exhausting all search options via other means (e.g., online searches, and requesting a copy of the study from the lead author directly), an inter-library loan was requested through the Health Library at Keele University.

3.3.5 Data Extraction

The next stage in the systematic review process involved development of an Excel standardized data extraction form, which was piloted prior to use on all included full texts.

Extensive data was collected on study characteristics, trajectory derivation analysis, trajectory-covariate association analysis (where applicable), and methodology used for trajectory analysis.

Thus, where possible, the following study level data was extracted: first author, publication year, country, study setting, data source (self-reported, routine/administrative records etc), type of data source (e.g., EHRs for UK data, Health Maintenance Organisations (HMO) data for US), sample size (both at baseline and follow-up in case of attrition), missing data, inclusion/exclusion criteria, participant characteristics, reason for work absence, definition of work absence (e.g., sickness absence only, or combination of sickness absence and disability pension), type of employment (self-employed, worker, employed), number of trajectory classes originally identified (if sequence analysis used), numbers of trajectory classes used in final analysis, name and description of each of the included trajectory classes, trajectory prevalence, summary of characteristics of individuals belonging to each trajectory classes, list of any confounders that were adjusted for when analysing trajectory-covariate associations, candidate covariates used in trajectory-covariate association analysis, any covariates that were categorised prior to analysis, specific definitions used for any of the covariates, which trajectory class was used as the reference group in trajectory-covariate association analysis, how many trajectory classes were used in total in the trajectory-covariate association analysis, summary of associations observed between covariates and identified trajectories (e.g. using relative risks), full description of the repeated longitudinal measure used for trajectory derivation, overall follow-up time and time intervals used in trajectory analysis, longitudinal trajectory analysis method used, statistical software used, summary of steps undertaken in generating trajectory classes (including any assumptions made), and summary of methodology used for trajectory-covariate association analysis. The data extraction form also allowed for any further useful observations to be recorded as additional notes.

Data extraction was performed by AL and GWJ, with any discrepancies resolved in consensus meetings.

3.3.6 Appraising Quality of Included Studies

In any systematic review it is important to address the risk of bias of included studies, this helps evaluate the validity and reliability of the study data being used in the synthesis stage, and hence ascertain whether the conclusions being drawn from the systematic review are indeed accurate. There are various tools that can be used to assess risk of bias, specific to different study designs. These risk of bias tools can be used, for example, to inform a sensitivity analysis whereby studies at higher risk of bias are removed and analysis is repeated using the remaining studies to see if this impacts the findings of the systematic review.

To assess the risk of bias for cohort studies in general, two of the main tools available are: the Newcastle-Ottawa Scale (NOS)[80] and the Critical Skills Appraisal Program (CASP) Cohort Study Checklist.[81] Other more specific risk of bias tools also exist for cohort studies, such as the Quality In Prognosis Studies (QUIPS) tool, which is used for assessing risk of bias in cohort studies that aim to identify prognostic factors.[82]

However, for cohort studies in general, the NOS tool is commonly used, and has been recommended over similar risk of bias tools in this context, particularly for its simplicity and ease-of-use.[83,84] After consulting with the Keele Systematic Review Team as well as the Keele Library Team, and considering the practicality of the NOS, it was decided to use this tool as a base in assessing risk of bias for this systematic review, with the possibility of adding elements from other risk of bias tools as necessary.

The NOS tool appraises the quality of a study using a series of questions relating to three broad perspectives: selection of cohorts, comparability of cohorts, and assessment of outcome – a template of this tool in its full original format is shown in Appendix B.

Firstly, presence of selection bias can result in cohort groups that are not representative of the target population, hence the findings of the study may not be generalizable, nor accurate. The NOS tool addresses this directly with questions asked in the template about: the representativeness of the exposed and non-exposed cohorts, and the ascertainment of the exposure – with clinical records and structured interviews being preferred. To gain a lower risk of selection bias with the NOS tool, it is also expected that study authors demonstrate that the outcome was not present at the start of the study.

For this systematic review, whilst the NOS tool was largely useful in its original form, not all of its elements were relevant. Hence use of an adapted risk of bias tool was necessary. Thus, the following adaptations were made to the selection bias elements of the NOS tool (with stars used to indicate responses with a lower risk of bias):

- The wording of question 1 was altered, to reflect the representativeness of the ‘study cohort’ (instead of 'exposed' or 'non exposed' groups):

  Representativeness of the study cohort (selection bias)

  a) truly representative *

  b) somewhat representative *

  c) selected group of users (e.g., nurses, volunteers)

  d) no description of the derivation of the study cohort
- Question 2 was deleted (as this relates to representativeness of the non-exposed group, which is irrelevant for this systematic review)

- The wording of question 3 was amended to reflect ascertainment of ‘work absence’ (instead of 'exposure'), and the wording of the preferred option of ‘secure record’ was amended to reflect what this means in the context of this systematic review – electronic health records:

  Ascertainment of work absence (measurement bias)

  a) electronic health records *

  b) structured interview *

  c) written self-report

  d) no description
- Question 4 was deleted (as this required demonstrating that the outcome of interest, sickness absence in this systematic review, was not present at baseline - but this review’s exclusion criteria already aimed to screen out studies that did not have a baseline sickness absent population, so inclusion of this question was redundant, as it would under-estimate true risk of bias by default)

It is also important that study findings are comparable, and as such the NOS tool expects that study analyses have been controlled for the most important factors that might affect the observed results. However, assessment of comparability of cohorts in the original NOS tool relates to exposed and non-exposed groups, which was irrelevant for this review, hence this NOS question was not used.

Furthermore, in trajectory analysis, typically the first step involves derivation of trajectories through fitting an unconditional model (i.e., without any covariates), then as a second step, a conditional model is fitted to test for the effect of any covariates on trajectory membership (this process is outlined in more detail in Chapter 5).[70,85,86] Therefore, assessing whether a study controlled for the effect of confounders is only

relevant for the second objective of this review. Two questions from the CASP tool concerning confounding in analysis were considered for use in this review (for the second objective only) but were ultimately not used, as few of the included studies performed the relevant multivariable analysis of trajectory-covariate associations.

Finally, the NOS assesses the risk of bias relating to the study outcome, this is achieved by first investigating the way in which the outcome is assessed – with blind assessments or record linkage preferred. In the context of this review, study outcome corresponds to trajectories of work absence, hence again, measuring work absence status during follow-up using EHRs is preferred over self-reported means. Then, a question is also asked to ensure that study follow up time is sufficient (spurious findings could otherwise have been observed) – in this review, best judgment by AL and GWJ was used, to consider the total follow-up time duration, and number of time intervals within that timeframe. Lastly, consideration is given to evaluating loss to follow up and the effect this may have on the validity of results. All three of these outcome related questions from the original NOS tool were retained.

Use of the QUIPS tool and remaining aspects of the CASP tool (aside from the confounding questions) were also considered, but these were deemed to be too detailed and complex for the needs of this systematic review. Moreover, several of the main aspects of these tools had already been covered by the above adapted risk of bias tool: study participation, study attrition, and outcome measurement.

The final version of the adapted risk of bias tool used in this systematic review contained five questions and is provided in full in Appendix C. A star system was used to rate risk of bias, with studies awarded a maximum of one star for each question (with higher stars indicating better quality studies).

Risk of bias was then graded into arbitrary categories assigned by AL:

- 0-1 stars indicated ‘high’ risk of bias,
- 2-3 stars ‘medium’ risk of bias, and
- 4-5 stars ‘low’ risk of bias.

To ensure that this adapted risk of bias tool was suitable for this systematic review, it was first piloted by AL, then quality appraisal was performed by both AL and GWJ, with discrepancies resolved in consensus meetings as required.

3.3.7 Analysis

The outcome of this systematic review was trajectories of work absence due to a MSK or MH condition. This outcome was derived using repeated measurements of work absence.

Typically, in a systematic review, if there is sufficient and suitable quantitative data, a meta-analysis is performed. A meta-analysis is a statistical analysis technique used to aggregate data from several individual studies that largely address the same research question with similar study populations, and allows a single pooled estimate of the outcome (such as treatment effect) to be calculated.[74] Hence a meta-analysis can be a powerful and useful tool, due to its ability to allow a large collection of results to be synthesised into just one result, that accounts for all of the information across all studies.[74]

In this review, an aggregated meta-analysis to synthesise prevalence of trajectory classes across included studies was planned to be conducted, if data were suitable (in terms of there being similar definitions of trajectory classes, and similar participant populations across the studies) and sufficient. The Cochrane Consumers and

Communication Review Group suggests that two studies can be considered as a sufficient number to perform a meta-analysis, but this group also emphasizes that this is only the case when these two studies can be meaningfully pooled, whereby comparable outcomes and participant populations are present in the studies.[87]

If the meta-analysis was possible, it was planned that summary estimates of pooled trajectory class prevalence would be presented, alongside accompanying 95% confidence intervals to display the level of uncertainty in the aggregated outcomes.

In addition to the principal planned meta-analysis, meta-regression analyses of the following subgroups were also planned:

- Health condition causing the baseline work absence (a MSK health condition alone, a MH condition alone, or a combination of MSK and MH conditions)

- Healthcare system (for example, comparing EHRs from UK data, against HMO data from the United States of America, and the Scandinavian healthcare system, amongst others)

However, a meta-analysis was ultimately not possible, as a small number of (heterogenous) studies were included in this review. Hence, to address the first and second study objectives (assessing for presence of derived trajectories, and trajectory-covariate associations based on multivariable analysis, respectively), AL performed a narrative synthesis using textual descriptions and tabulation of trajectory information.[88]

The third study objective, a critical appraisal of trajectory methods used, was also performed by AL, through a narrative synthesis.[88] Here the different methods used to model the trajectories were tabulated, described, and compared with respect to strengths

and weaknesses (these results are presented in a later Chapter on trajectory methods, in Section 5.4).

3.3.8 Systematic Review Update

As the original database searches in this review were conducted on 8th April 2021, it was deemed appropriate to perform a systematic review update to assess whether any new studies (that met the inclusion/exclusion criteria of this review), had since been published.

The search strategy from Section 3.3.1 was preserved, and searches were re-run, but with a date filter applied from 8th April 2021 to 17th July 2023. Where possible, creation date was used, or otherwise the next best date filter available, such as publication date.

The de-duplication process from Section 3.3.2 was also repeated, to retain only unique records within and across databases from the new search. However, prior to screening this new set of records, further de-duplication was performed against the unique records from the original search, using procedures set out in a study by Bramer et al.[89]

All remaining steps of the original systematic review were then repeated using this new set of de-duplicated records (title and abstract screening, full text screening, data extraction, quality appraisal, and analysis).

3.4 Results

3.4.1 Original Search Results

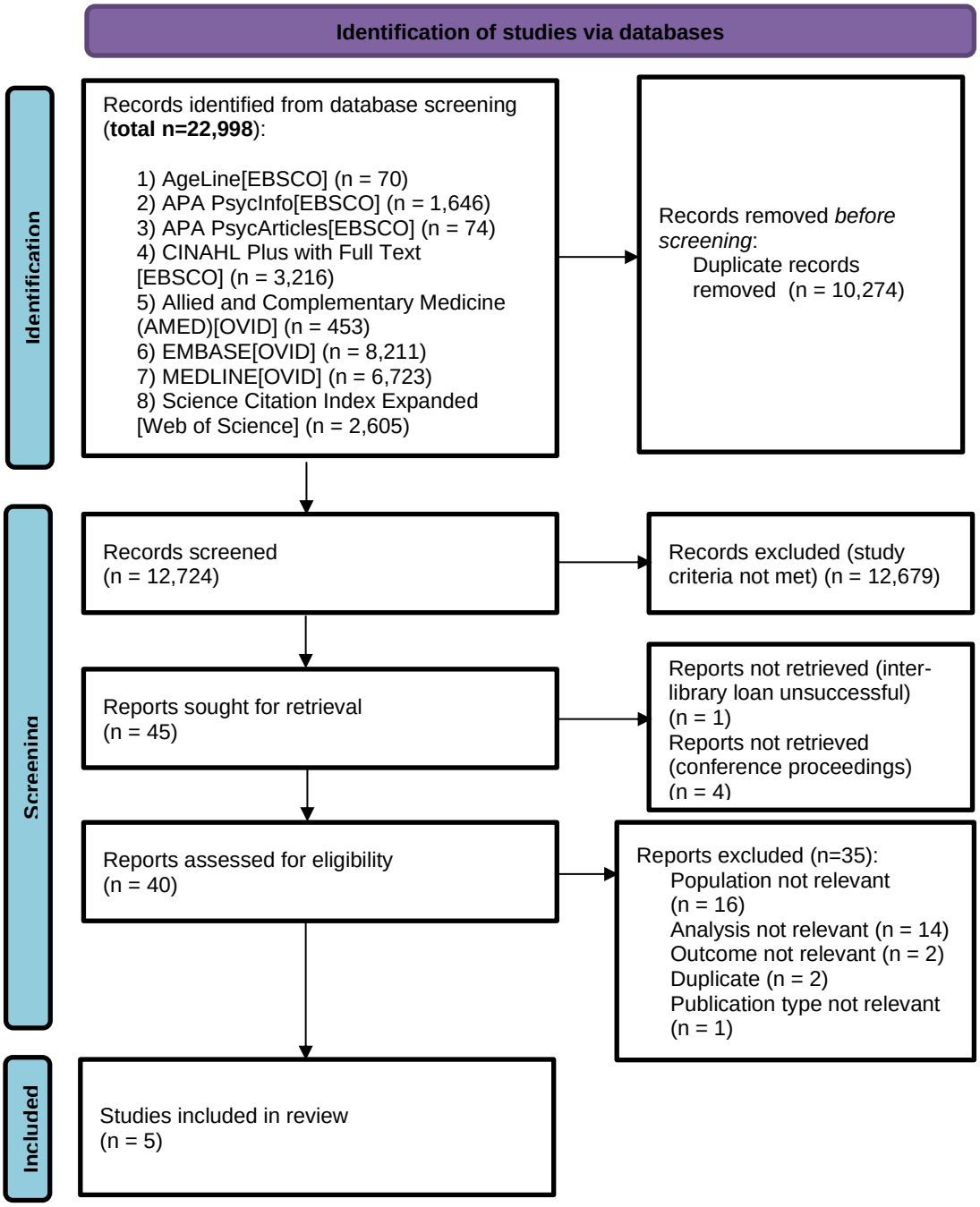
In this Section, screening results are presented first for the original search (up to 8th April 2021). The number of studies, at each stage of the searching and screening

processes are displayed in Figure 1, in accordance with the PRISMA 2020 flow diagram.[90]

After applying the bespoke search strategies to each database separately, a total of n=22,998 potentially relevant records were obtained. De-duplication processes in Mendeley Desktop removed n=10,274 of these, so that n=12,724 records were carried forward and imported into Rayyan software for screening. Title and abstract screening (as a combined step) then removed the majority of these, so that n=45 studies remained for full text screening. There were few conflicts (<1%) between AL and GWJ in the title and abstract (combined) screening process, these were easily resolved in a consensus meeting.

Of the remaining studies, n=4 were rejected due to the records being conference proceedings that had previously gone undetected as such in the screening processes.[91–94] Furthermore, n=5 inter-library loans were requested from the Health Library at Keele University;[95–99] these were for: n=4 studies which had a title available, but not an abstract (both AL and GWJ could not locate the abstracts after concerted efforts searching online);[96–99] and n=1 study where the full text could not be located (best possible efforts were first exhausted by searching online and asking for expertise from the Keele Library Team).[95]

Figure 1. Systematic Review PRISMA 2020 Flow Diagram for Original Search (up to 8th April 2021)



The inter-library loan came back unsuccessful for n=1 study,[98] with the librarian comment: 'unable to fulfil as no UK supplier found', hence this study was excluded from this review. Thus, in total n=40 studies progressed to full text screening stage.

After full text screening, n=35 studies were rejected, mostly due to the study population not meeting this review's inclusion criteria (n=16), or the type of analysis not being relevant to this review (n=14). There were some (n=10) conflicts between AL and GWJ in the full text screening process, these were easily resolved in a consensus meeting with all n=10 ultimately rejected.

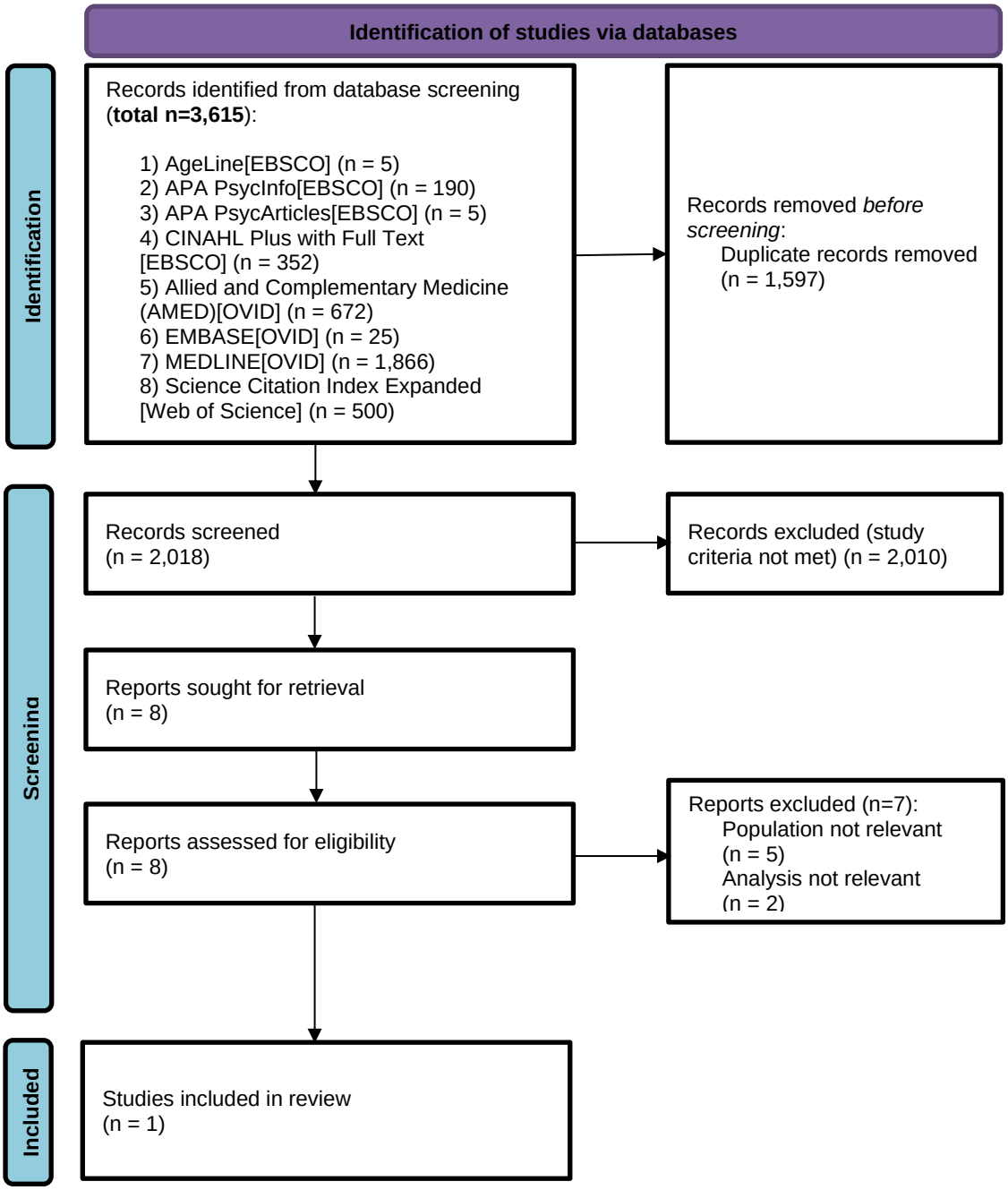
Hence, from the original search up to 8th April 2021, a set of n=5 studies remained after screening.[69,100,70,72,73]

3.4.2 Updated Search Results

Now, additional search results from the updated search are presented, covering the period from 8th April 2021 to 17th July 2023.

As shown in the PRISMA flow diagram in Figure 2, there were n=3,615 records after initial application of the search strategies. After de-duplication, both within these n=3,615 records, as well as through cross-referencing against the n=12,724 de-duplicated records carried through for screening from Section 3.4.1, n=2,018 records remained and were imported into Rayyan software for screening. Title and abstract screening (as a combined step) then removed the majority of these, so that n=8 studies remained for full text screening. There were no conflicts between AL and GWJ in the title and abstract (combined) screening process.

Figure 2. Systematic Review PRISMA 2020 Flow Diagram for Updated Search (searching from 8th April 2021 to 17th July 2023)



Full text screening removed n=7 studies, as before, mostly due to the study population not meeting this review's inclusion criteria (n=5), or the type of analysis not being relevant to this review (n=2). There were no conflicts between AL and GWJ in the full text screening process.

Hence, the updated search up to 17th July 2023 identified n=1 extra study after full text screening.[101] Thus, including the previously identified n=5 studies, there was a final set of n=6 studies potentially included for analysis in this review.[69,100,70,72,73,101]

3.4.3 Excluded Studies at Full Text Screening

Prior to describing the n=6 included studies in this review, the n=42 studies that were excluded after full text screening are now summarised. This is done to further explore the reasons for exclusion, especially as some of these studies presented trajectories of absence.

Firstly, n=2 studies were excluded for an absence definition that was not relevant to this review. Borg et al,[102] explored prediction of a disability pension state from a baseline sickness absence state, rather than prediction of RTW, nor were any trajectories were derived in this study. In contrast, Hakulinen et al[103] did derive trajectories, but these were based on repeated measures of employment (yes/no for employed/unemployed), rather than sickness absence.

The majority of the remaining excluded studies were due to the study population (n=21) or type of analysis (n=16) not being relevant to this review.

Of the n=21 studies excluded for reasons of the study population not being relevant, many were excluded due to not all of the population being sickness absent at baseline. For example, Ayala-Garcia et al[104] conducted a study on a Spanish population, and derived annual work absence trajectories due to a mental or behavioural disorder, separately for females and males. Then, subsequent trajectory-covariate association analysis was performed using multinomial logistic regression (this technique is described in more detail in a later Chapter concerning trajectory-covariate association

analysis, in Section 8.3). Both of these analyses were relevant for the first and second objectives of this review, yet this study was excluded as it was not clear which individuals (if any) had a baseline sickness absence, and there was no analysis present that isolated trajectories specific to this subgroup.

In another study, by Hou et al,[105] based in Taiwan, absence trajectories were also derived. However, whilst the study population was sickness absent at baseline, this study did not meet this review's inclusion criteria as it was based on individuals that were hospitalised with traumatic injury, not a primary care population.

Some examples of the n=16 studies excluded due to an analysis that was not relevant for this review, included studies that defined 'trajectories' a-priori, rather than by using the participant follow-up data. For example, Bültmann et al (2007)[106] derived RTW 'trajectories' as percentage changes in four RTW states between baseline (one month post MSK disorder) and end of follow-up; these 'trajectories' were not data driven. Given that this study was conducted whilst trajectory research was still in its infancy, especially with respect to work absence literature (for context, in this review itself, the earliest publication year of the six included studies was 2016), lack of knowledge and use of trajectory methodology could have limited the ability of the authors to apply such methods to address their research problem.

Other examples of studies rejected for identifying 'trajectories' a priori include the work of Baldwin et al (2006)[107] and Côté et al (2008).[108] These studies referred to the 'trajectories' as 'patterns' instead and were conducted over a similar time period to the Bültmann et al (2007)[106] study.

Finally, there were several instances where studies were rejected for not conducting a type of trajectory analysis, but instead performing other types of longitudinal data analyses, such as multi-state modelling, or a methodological study which used types of hazard functions to account for new and recurrent sickness absence.[109] Further exploration as to why trajectory methods are preferred over these other types of longitudinal analyses is provided in Section 5.1.1, in a later Chapter on trajectory methodology (Chapter 5).

3.4.4 Characteristics of Included Studies

The six studies included in this review were all published recently (2016-2023) and were diverse in their study populations (Table 4). None were conducted in the United Kingdom. There was one study from Canada,[70] and the rest were from Europe (two studies from Sweden,[69,72] and one each from: The Netherlands,[73] Denmark,[100] and Norway).[101]

Baseline sample sizes ranged from n=549,[101] to n=81,062.[70] Reason for work absence was due to various types of MH condition in n=3 studies,[100,72,73] and types of MSK condition in the rest.[69,70,101]

Finally, work absence was defined by sickness absence alone in n=4 studies,[100,70,73,101] and using a combined measure of both sickness absence and disability pension in the remaining two studies.[69,72]

Table 4. Summary of the Main Characteristics of Included Studies in Systematic Review

First Author, Publication Year	Country	Baseline Sample Size, n	Reason for Work Absence	How Was Work Absence Defined?
Pedersen, 2016	Denmark	725	MH Reason	Sickness Absence
Farrants, 2018	Sweden	10,327	Depression	Work Disability[a]
McLeod, 2018	Canada	81,062	MSK Disorders (Work Related)	Sickness Absence
Farrants, 2019	Sweden	4,894	Osteoarthritis	Work Disability[a]
Spronken, 2020	The Netherlands	9,517	MH Problems	Sickness Absence
Rysstad, 2023	Norway	549	MSK Disorders	Sickness Absence

Abbreviations: MH = Mental Health; MSK = Musculoskeletal
[a] Measured using sickness absence and disability pension

A more detailed summary of the study characteristics of included studies is presented in Table D.1 of Appendix D, covering a description of missing data (where applicable), inclusion and exclusion criteria, baseline participant characteristics, as well as how the reason for work absence was defined (for example, using particular ICD-9 or ICD-10 codes) and which work absence database was used.

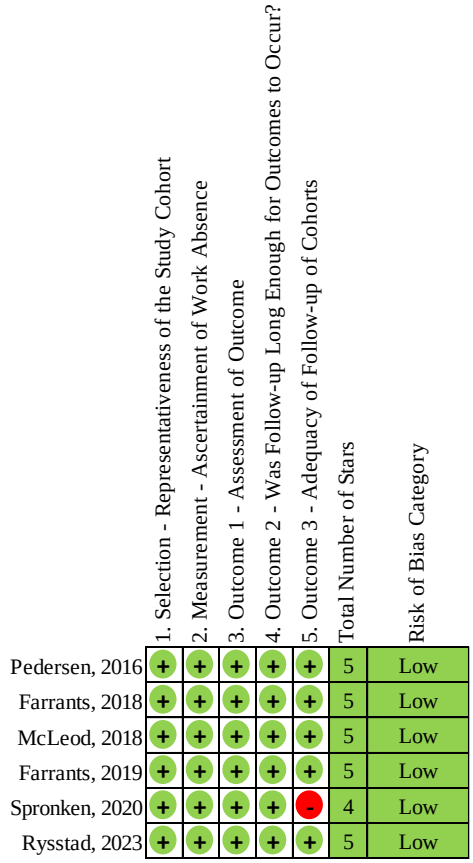
Appendix Table D.1 mostly highlights the diversity of the included studies. Albeit there were a few similarities in baseline characteristics amongst the studies. For example, three studies had inclusion criteria requiring individuals to be aged 16-64 years,[69,70,72] another study required ages 18-64 years,[100] and a fifth study 18-67 years.[101] Furthermore, the average age (expressed as either a mean or median) of individuals of four studies was between 41 to 42 years at baseline.[70,72,73,100]

3.4.5 Risk of Bias

The included studies in this systematic review were all generally of high quality. In Figure 3, the risk of bias is summarised by study and each question of the bespoke risk of bias tool used (as described earlier in Section 3.3.6).

All studies showed low risk of selection bias, as respective study cohorts were judged to be 'truly' or 'somewhat' representative of the target population in all cases. A clear strength of n=4 of the included studies in particular, was that they were population-based cohort studies with large sample sizes, hence no substantial issues of generalisability of findings were anticipated.[69,70,72,100]

Figure 3. Risk of Bias for the n=6 Studies Included in the Systematic Review



Note: Green circles represent instances where a star was awarded to a risk of bias question (indicating a lower risk of bias for that domain), whilst red circles represent where no star was awarded (indicating a higher risk of bias).

Similarly, all studies performed well in terms of measurement bias related to ascertainment of work absence – all used administrative records to determine work absence. However, a potential caveat is that there could be unidentifiable recording errors present in the administrative records (for example, recording errors when inputting the work absence data into the computer system).

Furthermore, there were no clear concerns with the quality of outcome data, with all six studies using record linkage to determine outcomes. Although, as with ascertainment of baseline work absence, there could be unidentifiable recording errors present. Additionally, it is possible that these records are not complete in their recording of RTW, which may also have gone undetected.

Follow-up time was either one or two years in all studies (and with either weekly or monthly time intervals used to derive trajectories), this was deemed long enough to allow outcomes to occur and capture sufficient variability in derived trajectories. Hence, all studies also received favourable ratings for bias relating to follow-up length.

Finally, five studies also scored well when assessing the adequacy of follow-up of cohorts. Three of these studies had complete follow-up of included study participants,[69,70,101] and the other two a small amount of loss to follow-up (only 0.6% of the baseline cohort was lost in one study, and this was due to death or emigration,[100] and 0.1% lost in another study due to death).[72] However, in the study by Spronken et al,[73] 16.1% of included participants were lost during the data cleaning phase (for various reasons, including not having complete absence follow-up data, as well as data entry issues such as where the end date of the absence was reported before the start date). As it was unclear how many of these participants were lost specifically as a result of

incomplete follow-up data rather than a data error, this study was penalised for this risk of bias domain.

In terms of overall risk of bias performance, all studies were graded as having 'low' risk of bias. A total of four stars was awarded to Spronken et al,[73] and the remaining studies all received the maximum rating of five stars.

3.4.6 Narrative Synthesis: Summary of Derived Trajectories by Study

Due to there being a low number of included studies (n=6), and high heterogeneity between them (in terms of study characteristics – as shown in Appendix Table D.1, as well as in the number and nature of derived trajectories – as will be described in this Section), a meta-analysis was not deemed appropriate, hence a narrative synthesis was performed instead.

In this Section, a descriptive outline of the derived absence trajectories is provided for each of the six studies in turn. Then in the next Section, an overall summary is presented, comparing the absence trajectories across the pool of included studies.

In order to understand the context in which these trajectories were constructed, the range of trajectory definitions used are shown in Table 5. One key difference was in the duration of sickness absence at baseline across the studies. The definition of the start of trajectory follow-up varied from onset of sickness absence,[70,73] to three weeks post the start of the sickness absence,[72,69] and to the date that a baseline questionnaire was issued (individuals had four to eight weeks of preceding sickness absence at questionnaire issuance date in one study,[100] and a median of 35.8 days of preceding absence in another).[101] Further details of these trajectory definitions are provided throughout this Section.

Table 5. Overview of Trajectory Definitions Used in the Included Studies of this Systematic Review

First Author, Publication Year	Work Absence Definition	Baseline Definition	Follow-up Duration	Time Measurements
Pedersen, 2016	Employment status	Time of baseline questionnaire issuance[a]	51 weeks	Weekly
Farrants, 2018	Number of net SA and/or DP days	Day 21 of the index SA spell	13 months	Monthly
McLeod, 2018	RTW status	Onset of SA[b]	1 year	4 week intervals
Farrants, 2019	Number of net SA and/or DP days	Day 21 of the index SA spell	13 months	Monthly
Spronken, 2020	% RTW[c]	Onset of SA[d]	Up to 2 Years[e]	Monthly
Rysstad, 2023	Number of SA days	Time of baseline questionnaire issuance[f]	1 year	Monthly

Abbreviations: DP = Disability Pension; RTW = Return to Work; Sickness Absence = SA
[a] Individuals had already been on sick leave for 4 to 8 weeks when baseline questionnaire was issued
[b] Defined as the first day of a work-related lost-time disability claim
[c] Return-to-work percentages were categorised as: 0%, 1%–19%, 20%–39%, 40%–59%, 60%–79%, 80%–99% and 100%.
[d] Whilst follow-up started from sickness absence onset, inclusion criteria required a minimum of 29 days of absence
[e] Censored after first full RTW (where 100% were contract hours worked)
[f] Individuals had already been on sick leave for a median of 35.8 days when baseline questionnaire was issued

The number of sickness absence trajectories identified, and their corresponding details (name, description, and prevalence) are displayed in Table 6. Furthermore, where reported, any descriptive summaries of trajectory characteristics are also discussed in this Section (and displayed in full in Appendix Table D.2), whilst a review of any trajectory-covariate associations reported from multivariable analysis (i.e. relevant for the second objective of this review) is reported in Section 3.4.8.

Table 6. Summary of Derived Work Absence Trajectories from Studies Included in Systematic Review

First Author, Publication Year	Number of Trajectories in Final Model	Trajectory Number	Trajectory Name and Description	Trajectory Prevalence, %
Pedersen, 2016	8	1	Sickness absence (almost 100% of individuals on sickness absence for approximately the first half of follow-up, before slowly reducing to around 50% by end of follow-up)	44.0
		2	Fast RTW (initial 100% sickness absence state decreased rapidly to 0% by around month 3, which was then sustained until end of follow-up. Individuals who exited the sickness absence state largely entered a work state).	21.9
		3	Slow RTW (similar to fast RTW trajectory, except that 100% sickness absence state was sustained for approximately first 3 months, before decreasing steadily to 0% by around month 7, which was sustained until end of follow-up)	14.4
		4	Sickness absence/temporary support (after around 4 months with approximately all individuals on a sickness absence, rapid decrease to 0% by around month 8, which was sustained until end of follow-up. Individuals who exited the sickness absence state largely entered a temporary support state)	5.4
		5	Temporary support (similar to trajectory class 4, except no sustained 100% sickness absence state in the initial months, instead there was an immediate rapid decrease to 0% of individuals on a sickness absence, which was reached by around month 4).	5.1
		6	Unemployment (moderately fast decrease to 0% sickness absence by around month 5, which then remained low until end of follow-up. Individuals exited sickness absence largely to enter an unemployment state).	4.4
		7	Permanent support (very similar to the unemployment trajectory, except individuals exiting sickness absence largely entered a permanent support state).	2.4

		8	Relapse (% of individuals on sickness absence decreased from 100% to around 20% after first 3 months, before rapidly increasing back to 100% by around month 8, and then slowly decreasing again)	2.4
Farrants, 2018	6	1	Decrease to 0 (monthly days of SA/DP) after 4 months	43.0
		2	Decrease to 0 after 9 months	22.0
		3	Constant high (at around 30 days of SA/DP per month)	11.0
		4	Decrease, then high increase (decrease from a high net days of SA/DP of around 22 per month at the start of follow-up, to around 14 by month 5, before increasing again)	9.0
		5	Slow decrease (started off with high net SA/DP days of 30 per month, but decreased steadily and continuously to around 3 days per month by end of follow-up)	9.0
		6	Decrease, then low increase (decreased from around 17 net days of SA/DP per month at the start, to around 1 day per month by month 4, which remained until around month 6, before a steady increase was observed until end of follow-up)	6.0
McLeod, 2018	9	1	Early-sustained RTW (reached a sustained state of RTW by the 1st month)	49.7
		2	Short-delayed RTW (reached a sustained state of RTW during months 2–6)	30.6
		3	Early NRTW (reached end state of NRTW within the first 6 months)	6.7
		4	Long-delayed RTW preceded by SA (reached a sustained state of RTW by months 7–13, with preceding events predominantly a SA state)	4.2
		5	Late NRTW (reached end state of NRTW by months 7–13)	3.1
		6	Constant SA (remained in SA state throughout follow-up)	3.0
		7	Deferred SA (reached a sustained state of SA anytime during months 2–13)	0.9
		8	Long-delayed RTW preceded by MRTW (reached a sustained state of RTW by months 7–13, with preceding events predominantly a MRTW state)	0.8
		9	Unclassifiable	1.1
Farrants, 2019	5	1	Fast decrease (had no/very little SA/DP days per month after 4 months of follow-up)	36.0
		2	Medium fast decrease (had no SA/DP days per month after 5 months of follow-up)	29.0

		3	Slow decrease (had no SA/DP days per month after 10 months of follow-up)	15.0
		4	Fluctuating (started off with around 20 SA/DP days in the first month, which then decreased over the first few months, before steadily increasing again from month 5)	12.0
		5	Late decrease (started off with high SA/DP days per month of >25, which was sustained for the first 9 months, before steadily decreasing to about 15 days per month at end of follow-up)	8.0
Spronken, 2020	5	1	Fast RTW with little chance of relapse (average of 136 days of sickness absence follow-up and 1.96 transitions before full RTW of 100% of contract hours achieved)	49.5
		2	Slow RTW with little chance of relapse (average of 402 days and 2.47 transitions before full RTW achieved)	20.8
		3	Fast RTW with considerable chance of relapse (average of 194 days and 3.07 transitions before full RTW achieved)	11.1
		4	Slow RTW with considerable chance of relapse (average of 419 days and 3.54 transitions before full RTW achieved)	9.5
		5	Very fast RTW with very small chance of relapse (average of 49 days and 1.00 transitions before full RTW achieved)	9.1
Rysstad, 2023	6	1	Fast decrease (rapid decrease to 0 sickness absence days 4 months from first assessment, then sustained RTW)	27.0
		2	Moderate decrease (slower decrease to approximately 0 sickness absence days by around month 8, then sustained RTW)	22.4
		3	Persistent high (stable and high number of sickness absence days throughout follow-up)	18.2
		4	Persistent moderate (stable and moderate number of sickness absence days throughout follow-up)	12.8
		5	Slow decrease (steady decrease to 0 sickness absence days at around month 11)	12.4
		6	U-shape (fast decrease in sickness absence days in first 4 months, followed by recurrence of absence from month 8 onwards)	7.3

Abbreviations: RTW = return to work; NRTW = non return to work; MRTW = modified return to work; SA = sickness absence; DP = disability pension.

Firstly, Pedersen et al[100] used a trajectory derivation method known as sequence analysis (a critical appraisal of the trajectory derivation methods of the six included studies is provided in Section 5.4 of a later Chapter on trajectory methodology). Briefly, sequence analysis involves identifying subgroups of individuals that share similar ordered occurrences of discrete states over time (i.e., sequences).

Pedersen et al[100] considered the following five employment states in their sequence analysis (with weekly repeated measurements used):

- Sickness absence
- Working (defined as weeks where no benefits were being received)
- Unemployment (weeks where unemployment benefits were being received)
- Temporary support (social benefits, not including sickness or unemployment benefits, were being given temporarily to encourage subsequent employment)
- Permanent support (social benefits were being given in a more permanent form, for example, due to early retirement, or to partially compensate wage due to reduced ability to work)

Whilst some individuals that were subsequently unemployed during follow-up were included here, this study was retained in this review because most of the trajectories presented showed only a small amount or no unemployment over follow-up.

Eight trajectories of absence due to a MH condition were identified by Pedersen et al,[100] over a 51-week follow-up period. The most favourable trajectory was named ‘fast RTW’ (21.9% prevalence). Here individuals rapidly and continuously moved out of an initial state of 100% sickness absence, and into a RTW state, with approximately all individuals achieving a RTW after around three months of follow-up, and thereafter

maintaining this state. The second most favourable trajectory ‘slow RTW’ (14.4%) was also similar, except slower, with most individuals remaining in an initial sickness absence state until around month three, and then sustained RTW achieved after around months six to seven.

One of the least favourable of Pedersen et al’s[100] derived trajectories was named ‘sickness absence’, and had the highest prevalence (44.0%). This trajectory was characterised by approximately all individuals maintaining a sickness absence state until around months six to seven of follow-up, after which a progressively increasing number of individuals started transitioning out to other states (mainly RTW) until end of follow-up (with around 50% of individuals still on a sickness absence after the full year of follow-up).

Pedersen et al[100] also derived five other trajectories, which occurred with low prevalence, ranging from 2.4 to 5.4%. Four of these trajectories concerned transitioning from an initial sickness absence to states of: temporary or permanent support (two trajectories were derived for the former and differentiated by the duration that the individuals remained in the initial sickness absence state), or unemployment (this trajectory, with 4.4% prevalence, was irrelevant for this review). The final trajectory was named ‘relapse’, whereby individuals moved out of the initial sickness absence state quickly (sickness absence decreased from 100% to approximately 20% by month 3), before rapidly returning to a sickness absence state (100% was reached by around month 8).

No form of descriptive participant characteristics were presented by Pedersen et al[100] for their eight identified trajectories (for the group of individuals sickness absent due to a MH condition).

Sequence analysis was also used in the study by McLeod et al,[70] now for absence due to a work-related MSK disorder, and with a trajectory definition based on four RTW states:

- Sickness absence
- Modified RTW (whereby an employee had returned to work but was working reduced hours, and/or with amended duties)
- RTW (where the employee had fully returned to work by resuming pre-injury work duties)
- Non RTW (where the employee wasn't working albeit being deemed fit for full duties, or had reached a plateau in their medical recovery and had a permanent functional impairment or vocational rehabilitation referral)

One of the major limitations of the McLeod et al[70] study however, despite performing well in risk of bias assessments and scoring the maximum rating of five stars (see Section 3.4.5), was that daily RTW status data was combined into four-week intervals over the one-year follow-up. This was done in order to prepare the data into an easy-to-use format for analysis, and a decision was made to only include the data of the last day of each four-week interval in analysis. This approach may have led to a substantial loss of information and may not accurately depict the work absence behaviour of individuals during the month.

Nine trajectories were uncovered over a one-year follow-up, and the two trajectories with the most favourable outcomes contained the majority of the analysis data (80.3%). These were comprised firstly, of the most ideal trajectory, named 'early-sustained RTW', whereby individuals had a rapid and sustained RTW by the end of the first month of follow-up (49.7% prevalence). The next best trajectory was named 'short-

delayed RTW', whereby full sustained RTW occurred by months two to six, after an initial delay due to sickness absence and/or modified RTW (30.6% prevalence).

The other seven trajectory prevalences were much lower, ranging from 0.8% to 6.7%, and constituted less favourable outcomes. Two of these trajectories did end in a sustained RTW during follow-up, but this occurred later than in the two most favourable trajectories, from months seven to twelve. A further two of these low-prevalence trajectories involved either a constant or 'deferred' sickness absence (other states such as modified RTW occurred before a sustained sickness absence was reached), and another two trajectories involved an end state of non RTW. Uniquely (compared to other included studies in this review), the final trajectory was as an 'unclassifiable' group (whereby the remaining 1.1% of individuals that could not be classified into any of the other eight trajectories were grouped together).

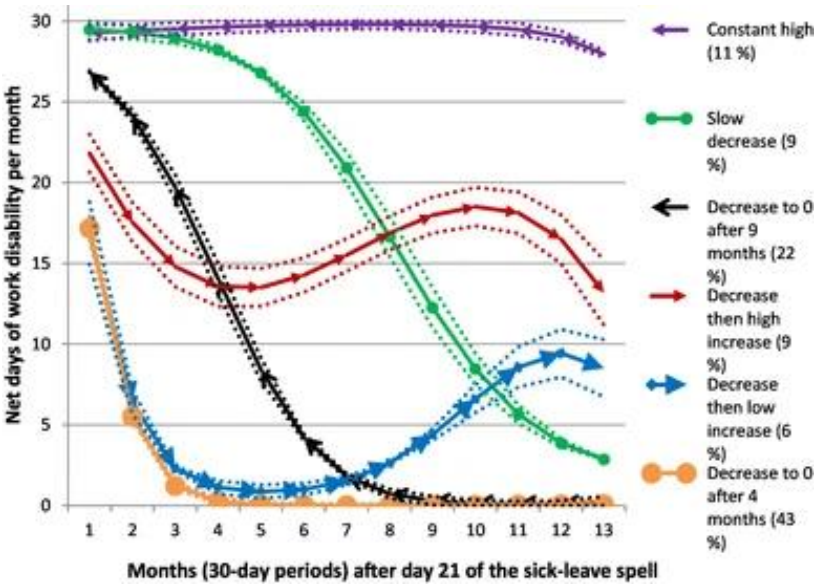
A descriptive summary of characteristics (see Appendix Table D.2 for full summary) showed that the most prevalent and favourable trajectory, 'early-sustained RTW', was characterised by a relatively higher percentage of younger individuals, compared to the other trajectories (for example, this trajectory had a prevalence of 49.7%, yet contained 60.1% of all 15-24 year olds from this study). Furthermore, there was a relatively higher percentage of lower extremity sprains and strains, or back sprains and strains, as the reason for sickness absence in this 'early-sustained RTW' trajectory. In contrast, one of the least favourable trajectories, 'constant SA', had a relatively higher percentage of older individuals, and torso fractures as reason for sickness absence.

Farrants et al (2018)[72] identified six trajectories based on absence due to depression. Unlike Pedersen et al[100] and McLeod et al,[70] Farrants et al (for both their 2018 study[72] and their 2019 study)[69] used a continuous definition of absence to derive trajectories,

and a form of LCA. Trajectories were computed using the mean number of net days of work disability per 30 day period, over a 13 month follow-up period. Work disability was defined as sickness absence and/or disability pension. Net days were used to allow for part-time absences. For example, two days of 50% absence were counted as one net day.

To provide a visual point of reference, Farrants et al's (2018)[72] trajectories are presented in Figure 4.[72(p682] Three of these trajectories showed clear continuous improvements in work absence behaviour over time (and were mainly differentiated by their relative speed to reach almost 0 work disability, as well as the initial levels of absence after one month of follow-up). The most favourable of these trajectories was named 'decrease to 0 after 4 months' and was also the most prevalent (43%).

Figure 4. The Six Work Absence Trajectories due to Depression Identified by Farrants et al (2018)[72]



Reprinted without any modification from "Work disability trajectories among individuals with a sick-leave spell due to depressive episode ≥ 21 days: A prospective cohort study with 13-month follow up," by K. Farrants, E. Friberg, S. Sjölund, and K. Alexanderson, 2018, J Occup Rehabil., 28(4), p. 682. Used under CC BY 4.0, available from DOI: 10.1007/s10926-017-9751-9.
Trajectory definition is work disability (sickness absence or disability pension), 95% confidence intervals of trajectories are shown.

Furthermore, there were two trajectories that presented fluctuating work absence behaviour over time, and a final trajectory with a constantly high work absence.

When comparing descriptive characteristics within these trajectories, Farrants et al (2018)[72] found that individuals in the most adverse trajectories ('constant high', and 'decrease then high increase') tended to be older, and to have had the highest proportion of sick leave in the year preceding the index sickness absence.

Then, in Farrants et al's[69] 2019 study (which was conducted with a similar structure to their 2018 counterpart,[72] albeit now with work absence due to osteoarthritis assessed instead of depression), five trajectories were identified. There were three trajectories of constantly decreasing work disability (again differentiated by speed to achieve RTW, as well as starting work disability level), one fluctuating trajectory, and one with constantly high work disability.

The most prevalent and favourable trajectory in particular, named 'fast decrease' (36%), was almost identical in shape and behaviour to the 'decrease to 0 after 4 months' Farrants et al (2018)[72] trajectory (which had a higher prevalence of 43%). Both trajectories began with an initial average of around 19 days of work disability after month one, which steadily decreased to almost 0 days after month four, and thereafter stayed at or close to 0.

Individuals belonging to the most adverse trajectories (in this case the only two trajectories with some sickness absence remaining at the end of follow-up - 'fluctuating' and 'late decrease'), were more likely to be older. In addition, individuals in these trajectories were more likely to be born outside of the European Union, and have more severe morbidity, in comparison to individuals in the other trajectories.

Uniquely in this review, in the study by Spronken et al,[73] a categorical definition of absence as a percentage of contract hours per month worked was used: 0%, 1%–19%, 20%–39%, 40%–59%, 60%–79%, 80%–99% and 100%. Relapse after a full RTW (whereby 100% of contract hours were worked in a given month) was not possible in this study, as full RTW was used as a trajectory end state. Trajectories were derived using a follow-up of up to two years (or, if earlier, time until full RTW). Whilst trajectories were measured from onset of SA, inclusion criteria of this study also required a minimum of 29 days of absence.

Five trajectories of absence due to MH problems were identified, and these were characterised by a combination of speed of RTW (fast or slow), as well as chance of relapse (high or low). Relapse was used to denote a deterioration in % of contract hours per month worked (but not from a full RTW).

The two most favourable trajectories derived by Spronken et al[73] both resulted in a 'fast' RTW, combined with little chance of relapse. The most favourable of these, 'very fast RTW with very small chance of relapse', resulted in the quickest time to a full RTW after an average of 49 follow-up days (9.1% prevalence). Whilst the second most favourable trajectory, 'fast RTW with little chance of relapse', resulted in full RTW after an average of 136 days, and was the most prevalent trajectory identified (49.5%).

A less favourable derived trajectory, 'slow RTW with little chance of relapse', exhibited a slower time to RTW (20.8% prevalence). Here an average of 402 days were needed until RTW.

Finally, Spronken et al's[73] remaining least favourable derived trajectories were characterised by a higher likelihood of relapse, and distinguished by the speed of RTW

('fast RTW with considerable chance of relapse' – 11.1%, and 'slow RTW with considerable chance of relapse' - 9.5%).

There was large individual variability when comparing descriptive characteristics of trajectories. Differences were observed with respect to gender, age, type of MH condition, organization sector and organization size between individuals in the slower and faster RTW trajectories, but no differences in terms of part-time compared to full-time employment.

In particular, an important finding was that women, older employees, and those working in non-profit sectors were more likely to experience longer RTW trajectories. Also, in relation to type of MH problem, the faster RTW trajectories were characterised by more individuals with stress complaints and adjustment disorders, whilst the slower RTW trajectories contained more individuals with burnout, mood disorders, and depression.

In the final included study of this narrative synthesis of derived trajectories, Rysstad et al[101] identified six sickness absence trajectories due to a MSK disorder over a one year follow-up. A continuous definition of monthly days of sickness absence was used, and this was then calibrated to a five-day working week, adjusting for the amount of sick leave and employment rate. Hence the maximum amount of monthly sickness absence possible was twenty days.

As seen in other studies of this review, Rysstad et al[101] also differentiated their (three) most favourable trajectories of decreasing sickness absence by speed of decrease (as fast, moderate, and slow). The most prevalent trajectory was also the most favourable, named 'fast decrease' (27.0% prevalence). Here, after initially starting off with approximately ten sick leave days by the end of the first month of follow-up (i.e., half of

the computed ‘working month’), sick leave decreased to approximately 0 days by the end of month four, and then remained at this level during the rest of follow-up.

The least favourable trajectories included two trajectories where sickness absence level was sustained over follow-up, either at a high level (‘persistent high’, 18.2%), or a moderate level (‘persistent moderate’, 12.8%). There was also a third less favourable trajectory named ‘U-shape’ (7.3%), which involved a relapsing back into sickness absence after an initial RTW.

Finally, when assessing descriptive characteristics of trajectories, in line with the other studies of this review, the least favourable trajectory of ‘persistent high’ had the highest median age (52.3 years). This trajectory also had the highest percentage of individuals who wanted a new job after the sick leave episode (36.4%). The highest proportion of women was observed in the other less favourable trajectories (74.3% and 75.0% in the ‘persistent moderate’ and ‘U-shape’ trajectories, respectively).

The ‘slow decrease’ trajectory had the lowest median age (47.1 years), and the most favourable trajectory, ‘fast decrease’ had the lowest median sickness absence days in the year prior to follow-up (30.0 days).

3.4.7 Narrative Synthesis: Summary of Derived Trajectories Across Studies

To conclude, in the narrative synthesis of derived trajectories in the previous Section, the trajectory definitions in the different studies of this review were first discussed. Various absence definitions were observed across the studies (employment and RTW states, % RTW in terms of contract hours per month, monthly work disability days, and monthly sickness absence days based on a work-week), with only the Farrants et al (2018)[72] and (2019)[69] studies using exactly the same absence definition (work

disability). Whilst similar follow-up times of approximately one year[69,70,72,100,101] were used (and one study with up to two years)[73], there were differences in the definition of baseline (with this either being onset of sickness absence, or a few weeks post onset).

Overall, a varying number of absence trajectories were derived:

- Spronken et al[73] and Farrants et al (2019)[69] derived five each
- Farrants et al (2018)[72] and Rysstad et al[101] derived six
- The two studies using sequence analysis, Pedersen et al[100] and McLeod et al[70], identified eight and nine trajectories, respectively

One commonality was that more favourable trajectories generally occurred with higher prevalence, whilst less favourable trajectories occurred with lower prevalence. Namely, for the McLeod et al[70] study, 80.3% of the trajectory prevalence was accounted for by the two trajectories which resulted in a sustained RTW from month six onwards. All of the trajectories derived by Spronken et al[73] in fact ended in RTW (and this was achieved by month fourteen), and 59.6% of trajectory prevalence was accounted for by the best two trajectories which resulted in the fastest sustained RTW without relapse. The three most favourable derived trajectories from Rysstad et al[101], of decreasing sickness absence, totalled 61.8% prevalence.

In the Farrants et al (2018)[72] study, the top two trajectories that resulted in no work disability from month nine onwards, accounted for 65% of the prevalence (rising to 74% if the next best trajectory is also included, that ended follow-up with an average of only around three days of work disability). The Farrants et al (2019)[69] study improved upon this further, with 80% of its trajectory prevalence occurring between the best three trajectories that resulted in no work disability from around months nine to ten onwards.

However, an exception was the Pedersen et al[100] study, whereby one of the less favourable trajectories of 'sickness absence' occurred with the highest prevalence (44%), and the two trajectories that resulted in sustained RTW only accounted for a combined 36.3% prevalence. The fact that Pedersen et al[100] considered three other states in their absence definition, in addition to RTW and sickness absence, as well as the relatively small sample size of n=725 may have contributed to these more unique findings.

Aside from fast and slow RTW trajectories, relapse trajectories were also observed across most studies.

Furthermore, Pedersen et al[100] and McLeod et al[70] each identified unique trajectories specific to the employment and RTW states, respectively, that they used to define absence (including trajectories characterised by temporary or permanent support employment states, and non- or modified-RTW states).

When comparing trajectory shapes across studies, there were several general similarities amongst the faster and slower RTW trajectories. For example, there were similar fast RTW trajectory shapes amongst all the studies:

- A fast RTW trajectory from Pedersen et al[100] (full RTW reached by approximately month three)
- McLeod et al[70] ('short-delayed RTW', whereby sustained RTW was achieved between months two to six)
- Farrants et al (2018)[72] (decrease to 0 SA/DP after four months),
- Farrants et al (2019)[69] (fast decrease of SA/DP trajectory, with approximately 0 SA/DP by month four onwards),

- Spronken et al[73] (‘fast RTW with little chance of relapse’ – average of 136 days to achievement of sustained RTW),
- and Rysstad et al[101] (‘fast decrease’, whereby sustained RTW was achieved at month four).

However, it is worth noting that prevalence’s of these mentioned trajectories varied considerably, from 21.9% up to 49.5%. Although, as shown earlier in Table 5, not all studies used the same definition of baseline, which does affect interpretation of these fast RTW trajectories.

Finally, where applicable, descriptive summaries of trajectory characteristics were also explored in the previous Section and were largely varied. Though, one commonality in the Farrants et al (2018),[72] Farrants et al (2019),[69] Spronken et al,[73] and Rysstad et al[101] studies was that individuals in the most adverse work absence trajectories tended to be older.

3.4.8 Narrative Synthesis of Trajectory-Covariate Associations

In relation to the second objective of this review, the McLeod et al[70] and Rysstad et al[101] studies were the only included studies that performed the desired multivariable analysis to evaluate trajectory-covariate associations. The other studies performed only descriptive comparisons of trajectory characteristics (as described earlier in Section 3.4.6), or a trajectory-covariate association analysis that was not relevant for this review (as explained in this Section).

McLeod et al[70] explored the effect of one covariate on trajectory membership - type of MSK disorder. Additionally, McLeod et al[70] opted to narrow down the nine previously identified trajectories to the most common six for this analysis (the ‘unclassifiable’

trajectory was removed, along with two other trajectories with <1% prevalence: 'deferred SA' and 'long delayed RTW-preceded by MRTW').

When comparing trajectory-covariate associations in the context of a finite number of trajectories, it is necessary to choose a referent trajectory to compare the other trajectories to. McLeod et al[70] chose their early-sustained RTW trajectory (the most favourable trajectory) as the referent group. Their analysis involved generating relative risk ratios with corresponding 95% confidence intervals (CIs), by using Poisson regression with robust standard errors. Furthermore, McLeod et al[70] carried out this analysis both unadjusted, as well as adjusted for covariates (age, gender, wage, firm size, prior claims history, occupation, and industry sector).

One of the key findings from McLeod et al[70] was that employees with back strains and sprains, compared to other types of MSK disorder, were generally most likely to follow the most favourable trajectory, early sustained RTW, than the other derived trajectories. Whilst employees with fractures or dislocations were more likely to follow sickness absence, more delayed RTW, or non-RTW trajectories, than an early sustained RTW trajectory. Adjusting analyses for covariates only slightly attenuated the estimated relative risks.

Rysstad et al[101] retained all six of their derived trajectories for covariate-trajectory association analysis. They used multivariable multinomial regression to test for association of nine covariates, decided a priori, with trajectory membership. The covariates used were: age, gender, education level, sick leave days in the prior year, RTW expectancy (self-reported by individuals), workability, pain intensity, multisite pain, and self-perceived health. Continuous variables were retained in their original format and not categorised prior to analysis. The only categorisation applied was to

‘education level’, to reduce the number of categories from four to two (‘low’ or ‘high’), in order to reduce the number of parameters being estimated in the model and remove sparse categories.

The least severe trajectory class, ‘fast decrease’, was chosen as the referent group. All covariates in the model were mutually adjusted (i.e., no extra variables were added to the model for adjustment), and (adjusted) odds ratios (OR) with 95% CIs were presented. Multicollinearity between covariates was also assessed and deemed as not present.

Key covariate-trajectory association findings from Rysstad et al[101] included that, compared to the ‘fast decrease’ trajectory:

- Lower RTW expectancy was associated with higher odds of belonging to two of the least favourable trajectories (‘high persisting’ and ‘moderate persisting’, with ORs of 1.39 and 1.32, respectively), as well as the ‘slow decrease’ trajectory (OR 1.18)
- Being female was associated with the ‘persistent moderate’ (OR 3.16) and ‘U-shape’ (OR 2.86) trajectories, but not the ‘persistent high’ trajectory
- Age, education level, and pain intensity were not associated with any trajectories

Pedersen et al[100] also performed a type of trajectory-covariate association analysis, by using logistic regression and applying covariate adjustments in the models. Though they tested association using the exposure group as the covariate (reason for sickness absence: a MH condition, compared to ‘other health reason’), and as there were no trajectory-covariate association analyses specifically for the subpopulation of

individuals sickness absent due to MH reasons, this was not valid for the purposes of this review.

The remaining three studies did not report any trajectory-covariate association analyses, but did perform the previously described descriptive analyses.[79,93,94]

3.5 Discussion

3.5.1 Key Findings

The first objective of this review, which was the main focus of this Chapter, was to determine published longitudinal trajectories of work absence in individuals with baseline work absence due to a MSK and/or MH condition. This was achieved through a comprehensive systematic review.

After initially identifying up to n=26,613 potentially relevant records from database searches, screening led to the inclusion of only six relevant studies in this review (published from 2016-2023), which emphasized the novelty of work absence trajectory research.

Five of these studies were conducted in Europe, and one in Canada, with none in the UK. Four of these studies were also large population-based cohort studies, with the highest sample size reaching n=81,062 individuals. Half of these studies looked at absence due to a MSK condition, and the other half absence due to a MH condition.

Considerable heterogeneity was observed amongst the included studies, with respect to the baseline study populations, study inclusion/exclusion criteria, the number of trajectories derived per study and corresponding trajectory prevalence's, as well as the participant characteristics of each trajectory.

For example, study populations differed in the amount of initial sickness already present at baseline (this varied from onset to a few weeks post onset of sickness absence). In addition, none of the six studies required a baseline incidence absence spell. McLeod et al[70] provided a broad descriptive summary of previous absence: 65% of their included participants had a previous sickness absence claim in the ten years prior to baseline. Then, Farrants et al in both their (2018)[72] and (2019)[69] studies presented a more detailed summary of prior sickness absence, where it was shown that individuals in the slower RTW trajectories generally had a greater amount of sickness absence days in the year prior to the index absence. This trend was also observed in the study by Rysstad et al.[101]

In contrast, data concerning sick leave prior to baseline were not available in the study by Spronken et al;[73] the study authors acknowledged this as a limitation of the dataset that they used. Pedersen[100] did not make any reference to previous sickness absence in their study, other than excluding individuals with more than three consecutive months of absence due to a MH condition in the year preceding the baseline questionnaire.

Thus, due to the heterogeneity and few included studies, the desired meta-analysis of trajectory prevalence's could not be performed. Instead, a narrative synthesis was conducted.

The narrative synthesis showed that whilst some differences in the trajectories were noted across the six studies, with certain study-specific trajectories identified, there were also similarities, especially relating to the faster and slower RTW trajectories.

In particular, a key finding was that more favourable sickness absence trajectories generally occurred with the greatest prevalence. These trajectories all involved a fast and sustained RTW, occurring within approximately the first five months from baseline.

For example, the most favourable trajectories identified by Farrants et al in their 2018[72] and 2019[69] studies, of a sustained RTW occurring within approximately four months from baseline (defined as 21 days post onset of incident sickness absence), were almost identical in shape and behaviour. The most favourable absence trajectories for the remaining four studies were also largely similar.[100,70,73,101]

Additionally, another key finding was that less favourable absence trajectories generally occurred with the least prevalence. Such trajectories typically involved a high and sustained level of work absence over follow-up, or a slow RTW (for example, in the study by Spronken et al,[73] the least favourable trajectory involved an average of 419 follow-up days until a 100% RTW of contract hours was achieved).

Furthermore, most of the included studies also uncovered trajectories involving relapses of absence.

Generally, from the pool of six included studies in this review, there was no discernible trend of any of the identified trajectories being a function of absence due to a MSK or MH condition. For example, as mentioned above, identical or similar trajectory shapes were observed for favourable trajectories relating to a fast RTW across all six studies; such trajectories were therefore not specific to reason for absence.

Comparing this review's MSK or MH condition absence trajectories against sickness absence trajectories present in studies excluded during the screening part of this review was more difficult, as most of these excluded studies did not have a baseline sickness absence population.

For example, Lalic et al[110] conducted a ten-year population study in Sweden, and absence trajectories were derived separately for two subgroups - those that had either a

strong or weak opioid administered for non-cancer pain in 2009. Baseline was defined as the date of opioid administration, and annual absence trajectories derived for a time span covering the five years preceding and succeeding baseline. As the focus was on sickness absence patterns before and after opioid initiation, the study population was not required to have had a baseline sickness absence. Unlike the included studies of this review, trajectories of decreasing absence occurred with low prevalence in Lalic et al's[110] study. Instead, Lalic et al's[110] most commonly occurring trajectories involved little or no absence during all of the 10 year follow-up (such trajectories were not possible by definition in this review's included studies). Additionally, Lalic et al[110] identified trajectories of increasing absence from baseline, these were also not applicable to this review's included studies.

However, an excluded study by Hou et al[105] was more comparable to our included studies, as individuals were baseline sickness absent here. Baseline was defined as the point of hospitalisation due to a traumatic limb injury (which led to this study's exclusion). Three trajectories were identified over a two-year follow-up period, using a binary definition of absence as RTW (yes/no) at each follow-up time point. The most prevalent trajectory (50.7%) was characterised by a favourable outcome of an increasing probability of RTW over time, with a sustained RTW achieved within six months of follow-up, whilst the least favourable 'slow RTW' trajectory was less prevalent (27.8%). These trends were consistent with the findings from our narrative synthesis. Additionally, the third identified trajectory further distinguished the speed of a fast RTW, with a sustained RTW occurring within one month of follow-up here (21.5% prevalence). In a similar manner, in this this review, Spronken et al[73] also differentiated trajectories between a 'fast' and 'very fast' RTW.

Descriptive statistics of the trajectories from the six included studies were also synthesised, and found to be varied, especially as a range of characteristics were explored across the studies. Nonetheless, one commonality was that individuals in less favourable absence trajectories were mostly older.

Finally, the second objective of this review assessed for presence of trajectory-covariate associations, derived using multivariable analysis. Only two studies, McLeod et al[70] and Rysstad et al[101] performed such analyses, both used their most favourable trajectory as the referent trajectory, but tested for the effect of different covariates. McLeod et al[70] only tested for the effect of type of MSK disorder on trajectory membership, and their key finding was that individuals with back strains were more likely to have a fast and sustained RTW, whilst those with fractures or dislocations were more likely to have a longer term absence trajectory or a non-RTW. Whilst Rysstad et al[101] explored a wider range of nine covariates, and a key finding was that individuals with self-reported negative RTW expectancy were more likely to follow trajectories of longer term absence. The excluded study by Hou et al[105], also showed agreement with this finding, with lower self-efficacy (in terms of a RTW occurring within one month here) associated with a slow RTW post traumatic limb injury.

3.5.2 Strengths and Limitations

One of the key strengths of this review is that it is believed to be the first systematic review investigating published trajectories of work absence. Although few relevant studies were ultimately identified, much useful information was gleaned about the types and prevalence's of work absence trajectories that do exist.

Furthermore, in order to identify as much relevant literature as possible, a comprehensive search was conducted, encompassing eight databases that were searched since inception up to 17th July 2023. Double screening was also performed throughout to ensure the integrity of the process to narrow down the potentially relevant records. Another strength was that the included studies were all graded as having a 'low' risk of bias.

However, one of the key limitations was that a meta-analysis of trajectory prevalence's was not possible, due to the lack of included studies and heterogeneity of included study data.

3.5.3 Conclusion

Overall, after a comprehensive search, only six studies (published from 2016-2023) were identified in this review of trajectories of work absence due to a MSK or MH condition, highlighting the scarcity of such research. Thus, work absence trajectory research has been identified as a relatively new and growing area.

Given the importance of trajectories in better understanding the potential future work absence behaviour of individuals who have an initial sickness absence, and the benefits that this knowledge could have in providing earlier and more targeted support to individuals at higher risk of sustained long-term sickness absence, more research on trajectories of work absence in necessary.

Furthermore, none of the included studies from this review were conducted in the UK. Absence management systems and primary care across different countries vary widely, therefore, in order to better understand the course of work absence in a UK setting, a study exploring trajectories of work absence in the UK would be beneficial.

In the next Chapter, several UK-based sickness absence datasets are contrasted for use in this thesis, and a final dataset is chosen.

Chapter 4. Dataset to be Used

The dataset to be used in all three studies of this thesis is described in this Chapter. First, different data collection methods are presented, followed by a comparison of suitable data sources.

4.1 Selection of Suitable Data Source

4.1.1 Data Collection Method

To achieve the main goal of this PhD and derive trajectories of work absence due to a MSK or MH condition and assess for the presence of characteristics associated with these trajectories, either primary or secondary data collection methods can be used.

If primary data collection is elected, this would first involve defining and then recruiting eligible individuals. Next, eligible individuals who consent to taking part in the study would be interviewed at baseline. Then, their work absence data over the follow-up period could be manually collected over repeated future time intervals (i.e., a prospective cohort study). Alternately, a retrospective study could be performed whereby individuals are asked to recall their work absence history.

Primary data collection generates self-reported work absence data, which is not always as reliable as its objective counterpart (electronically recorded fit note data), especially if individuals are being asked to recall their work absence history over a long timeframe (for example, over several months or years). However, recall bias can be limited if self-reported data is collected over short and recent time intervals. Wynne-Jones et al (2008)[111] performed a study which compared self-reported work absence to sickness absence certificates from EHRs. Self-reported sickness absence was assessed by asking individuals to state their number of days of sickness absence in the previous two weeks,

collected at baseline and at 12 months follow-up. The study showed an overall 95% match between the self-reported and EHR sickness absence data. Although, the self-reporting recall period in this study was conducted over a short time period (two weeks); this self-reported data would likely become more unreliable over long periods.

One of the other notable disadvantages of primary data collection methods is that such methods can be time-intensive and costly. The cost and time needed is generally increased by: longer study durations, larger numbers of study participants, and a greater amount of follow-up time points.

As mentioned in Section 1.4.1, in the UK, fit notes are recorded in primary care EHRs as routinely collected data. Thus, an alternate approach to access fit note data is through secondary data collection methods. Such methods allow fit note data to be analysed through a range of different electronic healthcare databases available for research from various organisations, subject to strict guidelines as to their use.

Advantages of EHR databases is that they often:

- Allow data to be accessed quicker than primary data collection
- Are rich in data (and thus are known as being sources of 'big data') – in terms of population coverage, length of follow-up (for example, some databases include follow-up of over 30 years)[60], and number of data variables available
- Are updated regularly
- Allow for greater generalisability in conclusions of research, due to the high population coverage (for example, the EHR for a particular country could be contained in a national database, whereas a study involving primary data collection may be focused on a specific local area of the country)

- Are cheaper to use than having to perform equivalent primary data collection
- Do not require such detailed and lengthy ethical approval processes as those required for primary data collection[112]

However, limitations of using EHR databases include that the data is not specifically collected for a research purpose, unlike a study based on primary data collection. Hence not all the specific data required by the researcher may be available with this approach (even though the general range of different variables is likely to be much greater through data linkages in electronic health databases).

Also, efforts may be necessitated in thoroughly checking, cleaning, and preparing the data for use before analysis – particularly owing to the vast size and complexity of some EHR databases. These additional tasks may require the researcher to undertake specialist training.

EHR fit note data may not be completely devoid of errors either, as it is dependent upon the data inputted by the user, as well as the computer system being used. Although, over the years, efforts have been made to improve EHR data quality, such as through data entry training for staff, as well as monetary incentives for high quality data reporting, as described further in Section 4.1.3.

In conclusion, and on balance, due to the numerous advantages of secondary data collection highlighted in this section (especially regarding speed of access, cost, generalisability of sample, and availability of large sample sizes with substantial follow-up – an important consideration when conducting longitudinal trajectory analysis), secondary data collection methods were used in this PhD.

In the next section, various electronic health databases in the UK are presented, and the chosen database is justified.

4.1.2 UK EHR Databases

In the UK, the largest research database of longitudinal medical records in primary care (and one of the largest such databases in the world), is the CPRD.[113]

Other similar, but smaller EHR databases also exist, as shown in Table 7. These include:

- The Secure Anonymised Information Linkage (SAIL) Databank,[114] a database specific to Wales;

- IQVIA Medical Research Data (IMRD),[115,116] formerly known as The Health Improvement Network (THIN) database,[117] a UK wide database with more than 20 million patients (of which around 4 million are currently registered); and

- QResearch, a UK wide database containing data for more than 35 million patients in total (including historical patients).[118]

In contrast to SAIL, CPRD contains data from all the UK devolved nations. Furthermore, CPRD is markedly greater in size than the datasets of SAIL, IMRD, and QResearch. CPRD contains data for more than 60 million patients in total (across its two databases: CPRD GOLD and CPRD Aurum), of which more than 18 million are currently registered patients.[60]

Table 7. Comparison of UK EHR Databases

Data Source	Coverage	Summary
CPRD	UK wide	- Contains >60 million patients (including >18 million currently registered) - Variety of data linkages available - In use for over 30 years and >3000 related publications
SAIL Databank	Wales	- Restricted to Welsh patients only - Established in 2007 - Variety of data linkages available
IMRD	UK wide	- Contains >20 million patients (approximately 4 million currently registered) - Over 1800 related publications
QResearch	UK wide	- Contains data for >35 million patients in total (including historical patients) - Variety of data linkages available
CiPCA	Regional	- Now defunct - Was specific to a subset of North Staffordshire general practices
Lambeth DataNet	Regional	- Specific to general practices in the London Borough of Lambeth

Abbreviations: CPRD = Clinical Practice Research Datalink; EHR = Electronic Health Records; SAIL = Secure Anonymised Information Linkage; IMRD = IQVIA Medical Research Data; CiPCA = Consultations in Primary Care Archive; UK = United Kingdom

A review comparing the growth and research outputs of the three foremost EHR databases in the UK (CPRD, IMRD, and QResearch), found that in the 10 year period from 2004-2013, CPRD represented the majority of the total publications arising from these three databases (63.6%), and also showed the strongest growth in publications, highlighting the more widespread use of CPRD.[119]

Since inception in 1987 up to February 2024, CPRD data has been used internationally across more than 3,000 peer-reviewed publications, covering a wide range of research areas, including: health care delivery, disease risk factors, drug safety investigations, effectiveness of health policy, and investigations of use of medicines.[120]

Additionally, a systematic review in 2010[121] found that CPRD, under its previous guise of the General Practice Research Database (GPRD), performed well when assessing the

validity of diagnoses recorded within it. For example, a median of 83.0% of cases due to the disease group MH and behavioural disorders were validated, as were a median of 80.0% of cases due to the disease group MSK system and connective tissue.

There are also smaller and more specific local EHR databases in the UK, including: Consultations in Primary Care Archive (CiPCA),[122] a Keele University database which was specific to a subset of North Staffordshire general practices (however, this database is defunct as of April 2022); and Lambeth DataNet,[123] a database specific to general practices in the London Borough of Lambeth.

For the analyses of this thesis, CPRD was chosen as the designated UK EHR database ahead of the other options, due to reasons of the:

- Larger population size it offers
- Reliability of its data (for example, with the high validity of the recording of MSK and MH condition diagnoses, as previously mentioned)
- Generalisability of the CPRD population compared to the general UK population (discussed in Sections 4.1.3 and 4.1.4).

Additionally, this decision was also guided by practicality, as Keele University already held a multi-study annual licence allowing access to CPRD data. Furthermore, the extensive CPRD experience present within the Keele University researcher community was also an influencing factor. A broad range of CPRD projects and successive publications have already been conducted in-house (with >80 Keele-based CPRD publications), such as the work by Rathod-Mistry et al (2021)[124] and Mason et al (2023).[125] Having easy access to an experienced pool of CPRD researchers was an

invaluable resource for any coding issues and general CPRD-related questions during data cleaning and analyses.

4.1.3 CPRD GOLD

The CPRD is sponsored by both the Medicines and Healthcare products Regulatory Agency (MHRA) and the National Institute for Health Research (NIHR) and is owned by the Department of Health and Social Care (DHSC). The origins of CPRD began with a small dataset known as Value Added Medical Products (VAMP), that was created in London in 1987, before growing and becoming the aforementioned GPRD in 1993, and eventually growing yet further into the CPRD in 2012.[113]

CPRD continues to grow today, with data collected from participating general practices on a daily basis, before being sent, anonymised, to CPRD on a monthly basis. These monthly builds of anonymised CPRD data are available for researchers to purchase upon approval of a data request protocol and payment. Furthermore, when requesting CPRD data for research, ethical approval is not required as the CPRD research database was already granted this approval from the Health Research Authority on 10th January 2022 (through the East Midlands - Derby Research Ethics Committee).[126]

The CPRD originally used only one database, CPRD GOLD, which is described in this section. The CPRD GOLD database uses one of the four main types of GP IT systems in England, Vision. A cross sectional study investigating the spatial distribution of the main GP IT systems in England showed that in 2016 only 9% of the included 7526 practices used Vision software, whilst the majority (56%) of practices used Egton Medical Information Systems (EMIS) software.[127]

In October 2017, CPRD introduced a second database, known as CPRD Aurum, that incorporated this more popular EMIS GP software.[128] Due to CPRD GOLD and Aurum using different GP IT systems, these databases differ in structure of the included data, and hence CPRD currently does not intend to merge them and instead offers them as separate databases.[129] Thus when requesting access to CPRD EHR data, a decision regarding use of CPRD GOLD or CPRD Aurum is required. CPRD Aurum is presented in the next Section and contrasted to CPRD GOLD.

The daily primary care EHR data that is used to generate the monthly CPRD GOLD builds is inputted by general practice staff, such as a GP or nurse. Originally, version 2 Read codes were used for this purpose.[113] Read codes are a hierarchical classification system consisting of over 96,000 codes.[113] General practice staff are trained in how to use Read codes, and these can be used, for example, to describe a patient's condition during a consultation (e.g., if a diagnosis is issued to a patient, this is recorded electronically as a Read code and linked to the consultation date).

However, an alternative coding system, SNOMED CT, was introduced in primary care in England in April 2018, and replaced Read codes.[130] The roll out occurred under a phased approach.[130] The EHR data in CPRD GOLD is coded using a combination of SNOMED CT and the aforementioned version 2 Read codes.

Whilst SNOMED CT codes are also a means of coding EHR according to a specific structure, they offer advantages such as: allowing key information to be shared in a uniform manner between different health and care settings; a greater range of possible details to be inputted by clinical professionals; inclusion of diagnosis and procedures, symptoms, family history, allergies, assessment tools, observations, devices; a greater

support for clinical decision making; easier clinical auditing; and less chance of the coding being interpreted incorrectly across different clinical settings.[130]

Some numerical data can also be recorded directly (without the need for a Read/SNOMED code), such as height and weight, or alcohol intake, but this data may be missing if the staff member did not take this measurement during a given consultation. However, good quality reporting is generally encouraged by means of the Quality and Outcomes Framework, which offers monetary incentives to GPs who record key data items.[131] Finally, GPs are also free to enter any uncoded free text notes about their patients, but this data is not available to researchers, particularly because it can often contain sensitive patient identifying information.

One of the other main advantages of CPRD data that has not been previously mentioned, is its incorporation of linked data. Linked data can be requested when submitting a data request protocol to CPRD and allows for the research data to be enhanced beyond primary care data, to span areas including: COVID-19 data, patient and/or general practice neighbourhood deprivation measures, hospital episode statistics, mortality data, and cancer registry data. The same linked datasets are available across both CPRD GOLD and Aurum, and there are plans for CPRD to further develop available linked datasets, to advance the reach of future research.

In the December 2023 CPRD GOLD monthly release (https://doi.org/10.48329/30pm-xq61),[132] there were:

- 21.4 million research acceptable patients, based on CPRD's bespoke metric for assessing the research quality of the contributing data of patients (this figure

includes patients that were initially registered to a practice included in CPRD, but then died or transferred out of their practice)

- 3.0 million patients (of the total 21.4 million) that were currently registered in UK practices (this corresponded to 4.4% of the UK population)
- 9.3 million patients (of the total 21.4 million) that had data that was eligible for linkage
- 984 total practices in the CPRD GOLD database
- 366 practices (of the 984) that were currently contributing data (representing 4.6% of the total UK general practices)

The n=366 currently contributing practices were spatially distributed across all four devolved nations, but with the least (1.9%) in England, and the highest in Scotland (56.8%).

In terms of follow-up time, the median was 5.6 years for the group of total patients, and 12.7 years for the group of currently registered patients. Thus, in both cases, lengthy follow-up data was available.

Since the introduction of EMIS software, there has been a substantial decrease in practices registered with CPRD GOLD (as evidenced by the large difference in total and currently contributing practices), especially in England. This is also evident in the large difference between the 21.4 million research acceptable patients, and the 3.0 million currently registered patients. This difference is accounted by the transfer of practices from CPRD GOLD to CPRD Aurum (which does support EMIS software), as well as patients that are now deceased, and patients that have transferred out of their practice otherwise (i.e., not to a CPRD Aurum practice).

In terms of generalisability, Herrett et al (2015)[113] showed in their study, that CPRD GOLD patients were largely representative of the UK census population from 2011, with respect to age, sex, and ethnicity.

4.1.4 CPRD Aurum

In this Section, the alternative CPRD database, CPRD Aurum is presented, and then contrasted against CPRD GOLD.

CPRD Aurum generally functions in a similar way to CPRD GOLD. The same primary care data is collected, relating to: diagnoses, symptoms, prescriptions, referrals, and tests. Linkages to equivalent secondary data sources are also permitted. However, as mentioned in the previous Section, the key difference is that the two databases use different GP IT software systems (CPRD Aurum uses EMIS, whilst CPRD GOLD uses Vision).

In the December 2023 CPRD Aurum monthly release (https://doi.org/10.48329/7njs-8a57),[133] there were:

- 46.6 million research acceptable patients (including patients not currently registered)
- 16.0 million patients (of the total 46.6 million) that were currently registered in UK practices (this corresponded to 23.9% of the UK population)
- 35.3 million patients (of the total 46.6 million) that had data that was eligible for linkage
- 1,771 total practices in the CPRD Aurum database
- 1,589 practices (of the 1,771) that were currently contributing data (representing 19.8% of the total UK general practices)

The n=1,589 currently contributing practices were now spatially distributed only in England (100%).

In terms of follow-up time, the median was 5.2 years for the group of total patients, and 9.5 years for the group of currently registered patients. Thus, in both cases, lengthy follow-up data was available.

As shown by Wolf et al,[128] and similar to CPRD GOLD, CPRD Aurum was also shown to be representative of the general English population in terms of age and gender, as well as deprivation and geographical spread (comparing a mid-2017 snapshot of CPRD Aurum data to mid-2017 data on the broader English population from the ONS).

Thus, although there are many similarities between CPRD GOLD and Aurum, the clear difference is the substantially higher patient population in CPRD Aurum (that results from the difference in GP IT software system used). For example, there were more than twice the number of total research acceptable patients in CPRD Aurum database, compared to CPRD GOLD, from the December 2023 monthly release. This difference then increased to more than fivefold when comparing the number of currently registered patients available in each database.

Another important difference is that CPRD Aurum is predominantly centred around only general practices in England, whereas CPRD GOLD has a low percentage from England. For example, there was 100% of the coverage of CPRD Aurum in the December 2023 build was from English practices (compared to 1.9% in CPRD GOLD), although CPRD does intend to start adding more general practice data from the other devolved nations to CPRD Aurum in the future.

Hence for reasons of vastly greater data availability, in terms of patient population size, and considering the many similarities between CPRD GOLD and Aurum otherwise, CPRD Aurum was the chosen EHR database in this thesis.

In the next section, the process to retrieve this data is specified.

4.2 Data Request and Protocol Approval

In order to access data from CPRD, there is a stringent data request process (the process is the same for accessing CPRD GOLD or Aurum), whereby approval of a protocol signed off by the CPRD's Research Data Governance (RDG) Process is necessary. Prior to the RDG process, it was necessary to gain approval from the Independent Scientific Advisory Committee (ISAC) for MHRA database research. However, following the recommendations of an internal review, and in an effort to ensure the future sustainability and adeptness of CPRD's internal data governance framework, the RDG process replaced the ISAC on 1st June 2021.[134,135]

Data request protocols are submitted to CPRD via the electronic Research Applications Portal (eRAP) online system. Then, the stages of the RDG process involve:

- A screening of potential applicants and funders, to ensure that the sensitive data being requested is only accessed by legitimate researchers and honourable organisations.
- This is followed by a brief triage assessment of applications for completeness and categorisation into routine or non-routine research.
- Research protocols classified as 'routine' are thoroughly reviewed by CPRD researchers (for feasibility of research plans and proposed methodology, and to

ensure potential research is of public benefit), whereas ‘non-routine’ protocols are reviewed by an Expert Review Committee (ERC).

- Both ‘routine’ and ‘non routine’ protocols may also be referred for further review to the Central Advisory Committee (CAC). The CAC also perform general quality assurance of the entire research governance process.
- Finally, a decision is made regarding the data request protocol (taking into account feedback from CPRD researchers and ERC) and communicated back to the original applicant. Protocol revision may be recommended before the data request can be accepted.

The data request protocol for research relating to this PhD was submitted on 17th November 2021, and subsequently approved without revisions on 22nd December 2021 (CPRD study reference 21_000665), and is provided in full in Appendix E.

As the original research in this PhD involved observational studies of routinely collected anonymised data (through the CPRD), and CPRD already has its own ethical approval through the Health Research Authority (see Section 4.1.3), no further ethical approval was required.

4.3 Protocol Amendment

Whilst performing study analyses, several amendments to the original approved CPRD protocol were deemed necessary. These amendments are explained in full in Chapter 6, Chapter 7 and Chapter 8.

A summary of the changes is that:

1) Only individuals that had a MSK or MH consultation in the two weeks prior to their index fit note were included (not ± two weeks as previously stated; discussed further in Section 6.2.2)
2) A random sample of individuals for trajectory derivation analyses was no longer taken, rather, all individuals that met the study population criteria were now included (Section 7.2.2)
3) Minimum required LCA trajectory prevalence was changed from 5% to 1% (Section 5.2.6)
4) Injury/poisoning and respiratory conditions were removed as specific comorbidities when testing for association with trajectories (Section 8.2.5)

The amended data request protocol was submitted to CPRD on 12/09/2023 and approved on 20/09/2023 (provided in full in Appendix F).

4.4 CPRD Aurum Data Used

The build of CPRD Aurum data that was used in this thesis, was the February 2022 release (https://doi.org/10.48329/gcgx-f815).[136] This database contained:

- 40.9 million research acceptable patients (including patients not currently registered)
- 13.4 million patients (of the total 40.9 million) that were currently registered in UK practices (this corresponded to 19.9% of the UK population)
- 37.5 million patients (of the total 40.9 million) that had data that was eligible for linkage
- 1,489 total practices in the CPRD Aurum database

- 1,358 practices (of the 1,489) that were currently contributing data (representing 16.6% of the total UK general practices)

The n=1,358 currently contributing practices were mainly spatially distributed in England (99.0%), with the remaining 1.0% in Northern Ireland.

In terms of follow-up time, the median was 4.8 years for the group of total patients, and 8.7 years for the group of currently registered patients. Thus, in both cases, lengthy follow-up data was available.

Furthermore, linked data was also requested (at patient level), for a deprivation measure. This is deprivation measure is described further in Section 8.2.1.

4.5 Conclusion

Now, with CPRD Aurum chosen as the final dataset for this thesis (for studies 1 to 3, in Chapters 6 to 8, respectively), and the build of the dataset summarised, it remains to discuss the statistical methodology that is used throughout this thesis.

For study 1 (Chapter 6), the analysis concerns incidence rates, which are relatively rigid in their application, without many choices to consider when performing the analysis. Hence the methodology used in Chapter 6 is explained within the Chapter itself.

However, studies 2 and 3 (Chapter 7 and Chapter 8) are based on trajectories of work absence, which is a novel and complex type of longitudinal statistical analysis. When performing trajectory derivation, there are several different methodological approaches that can be used, and a range of decisions to be made when applying each one – these different trajectory derivation methods are the focus of the next Chapter.

Chapter 5. Trajectory Derivation Methods

In this Chapter, an overview of different individual-centred approaches that were considered to model trajectories of work absence in this thesis is provided. This Chapter presents a technical statistical extension to the brief and more general discussion of the importance of trajectories in this PhD discussed in Section 1.4 of the Background Chapter.

Section 5.1.1, explains why using growth curve models (GCM) and taking an individual-centred approach is considered an appropriate way to analyse the longitudinal sickness absence data relevant to this thesis. In the remainder of Section 5.1, the foundation of LCA models and the first of the three chosen types of trajectory derivation model is presented: Latent Growth Curve Modelling (LGCM).

Following this, in Section 5.2, extensions of LGCM to two further chosen trajectory derivation methods are presented. These more complex types of LCA are: Latent Class Growth Analysis (LCGA) and Growth Mixture Modelling (GMM).

Then, other individual-centred approaches to longitudinal modelling are also introduced in Section 5.3; these approaches are then referred to in Section 5.4, whereby a critique of the different trajectory derivation methodology used in the six included studies of the systematic review in Chapter 3 is performed.

Finally, in Section 5.5, this Chapter closes with a discussion of the trajectory derivation methods that were ultimately taken forward for use in this thesis.

5.1 Latent Growth Curve Modelling (LGCM)

GCMs (often referred to interchangeably as trajectories) are a novel type of model, used to estimate patterns of change over time, and take an individual-centred approach. The focus of this Section is on LGCMs, which are the simplest type of GCM and involve fitting one common trajectory for an entire study population. In Section 5.2, extensions to the GCM are explored that allow for multiple different trajectories to be fitted for a study population (for example, through LCGA or GMM).

5.1.1 Why Use a Growth Curve Model (GCM)?

Before describing how to apply and interpret a LGCM, the rationale for choosing to use GCMs in this thesis (either as a single GCM or multiple GCMs for a study population), as compared to other analysis methods, is first discussed.

Most of the existing sickness absence literature for individuals with a baseline absence involves use of a dichotomous absence measure and analysis through either a cross-sectional approach based on a single time point (such as logistic regression), or a time-to-event approach (such as Cox regression).[69,101]

However, RTW is a complex and dynamic process that changes with time, and both the cross-sectional and time-to-event approaches may be sub-optimal in capturing this as they treat RTW as a fixed status. Much information regarding the continuity in change of the absence over follow-up is lost with these approaches. Indeed, time-to-event analysis has a different emphasis to what is required in this thesis, as the focus is on analysing the duration of time until an event occurs (such as a RTW), rather than change in patterns of absence behaviour over time.

A cross-sectional approach is limited too for the requirements of this thesis. Consider a study that uses logistic regression to evaluate absence (yes/no) after six months of follow-up since initial absence onset. This cross-sectional approach does not take duration of absence into account. For example, two individuals who are both on a sickness absence after six months of follow-up, one who had six months of continuous absence, and another with intermittent RTW spells and relapses back into absence throughout the six months, would both be treated the same and classed as 'absent' after six months. Furthermore, speed of RTW is not taken into account either. For example, an individual who has a fast and sustained RTW after one month of follow-up would be treated the same as an individual who experiences a slower RTW that occurs after six months of continued absence.

In contrast, using a GCM allows for repeated measures of absence data to be used to assign individuals into common subgroups (trajectories) based on heterogeneity in both speed and duration of RTW spells. In other words, a GCM allows RTW to be modelled as a process that changes over time.

Furthermore, a GCM is constructed at the individual level. That is, the within-individual development of a repeated absence measure (intra-individual changes) is first analysed, then assignment to common subgroups is conducted based on similarity in absence behaviour between individuals (inter-individual differences). For example, intra-individual development may identify a subgroup of individuals in a study who each have a similar trajectory of sustained absence behaviour over time, and these individuals could then be pooled together into one common sustained absence trajectory subgroup (due to their small inter-individual trajectory differences). Whilst there might also be another subgroup of individuals in the same study, each with a similar fast and sustained

RTW trajectory, leading to the derivation of a common fast RTW trajectory. This focus of GCMs in deriving subgroups based on inter-individual differences of intra-individual change is key, as in this thesis the aim is to better understand how to help people at the individual level to reduce their risk of being on a long-term sickness absence.

In contrast, population average models are another approach to analysing repeated measures of absence. However, in these models the focus is on the population level, not the individual level. Thus, any conclusions drawn from population average models are representative of developmental changes for the study population as a whole and not for specific individuals. These models take a variable-centred, rather than individual-centred approach to change over time.

For example, Karlson et al[137] used a type of population average model, generalized estimating equations (GEE), to assess long term stability of RTW for individuals on sick leave due to burnout, comparing a group of individuals receiving a treatment against a control group. Through use of GEE, and with considerable time points included (sick leave data was calculated every tenth week over 130 weeks), RTW was treated as a dynamic process and changes over time were assessed in terms of a binary repeated measure (a partial/full RTW of 25% or more = yes, compared to non-RTW). However, this study's focus was variable-centred and compared levels of the study population (split by a treatment and control group) that had a partial/full RTW (yes compared to non-RTW) at each follow-up time point, rather than tracking these changes in RTW development at the individual-level, as a GCM is designed for.

To conclude, GCMs were used in this thesis as they are suited for modelling development of absence over time at the individual-level, which is what is desired in this thesis to help better identify individuals at risk of a long-term sickness absence.

The next Section compares two different frameworks for fitting a GCM.

5.1.2 Choosing a GCM Framework

GCMs are typically fit using one of two frameworks, either a multilevel or a multivariate approach to the repeated outcomes.[138]

The multilevel framework incorporates hierarchical linear regression modelling. Repeated time measures are defined as independent variables at the lowest level ("level 1") and nested within individuals ("level 2"). This hierarchy can be further extended if desired. For example, individuals may be nested within a GP practice ("level 3"), and GP practices within a geographical region of England ("level 4"). In this multilevel framework, inter-individual change of the repeated measure over time is accounted for by random effects.

Alternatively, the multivariate approach is based on structural equation modelling (SEM). Specifically, a LGCM specification of SEM is applied, whereby change over time in the variable of interest is regarded as an unobserved, "latent" process. The latent variables for a linear LGCM (this model is discussed in detail in the next Section, 5.1.3) are the intercept (the initial value of the repeated measure), and the slope, which describes how this repeated measure changes over time. These two latent variables (through their mean and variance-covariance structure parameters) are analogous to the random effects from the multilevel framework and are used to describe inter-individual differences in growth curves.

Hence, it is possible to specify equivalent linear GCMs across these two frameworks when using random effects from the multilevel framework that correspond to latent parameters in the LGCM. Stoel et al (2003) showed in their study, that the parameter

estimates of a linear GCM of language acquisition in children during primary school (based on four repeated time measurements and a two-level hierarchy of time points nested within individuals), were indeed identical across both frameworks.[139]

However, the main difference between these two frameworks relates to the handling of the time variable. In the multilevel GCM framework, multiple time points are treated as observations derived from the same variable, which makes this a univariate approach (with respect to the GCM). Whereas, with the LGCM framework, each time point is treated as its own separate variable, which makes this a multivariate approach.

These differences in the handling of time, have been shown to make the LGCM framework more flexible than the multilevel approach, especially with regard to application of more complex GCMs.[139] For example, the following model extensions can be more easily implemented with the LGCM approach: incorporation of a higher-order growth model that contains multiple different types of repeated measures rather than a single repeated measure, and models that use both categorical and continuous latent variables to define the intra-individual heterogeneity in growth curves.[138,139] In contrast, the multilevel approach is more appropriate if a higher level of nesting beyond the individual level is required.[139]

Due to the greater flexibility that the LGCM framework offers, this framework was used throughout this thesis. In particular, through permitting use of both categorical and continuous latent variables, the LGCM framework is more suited to performing analyses based on finite mixture models (described in Section 5.2), whereby multiple different growth curves can be estimated for subgroups of the same study population, which is important to address this thesis' aims.

5.1.3 Linear LGCM with Continuous Repeated Measures

In this Section, the conventional LGCM based on repeated continuous measures over time is introduced. Furthermore, the focus of this Section is on an unconditional linear LGCM, the most basic form of a LGCM.

A path diagram is the traditional way to visually display a SEM. The corresponding path diagram for an unconditional linear LGCM based on a continuous repeated measure is shown in Figure 5, and the accompanying model equations in Box 1.

Classical SEM notation is used in this path diagram,[140] whereby latent (unobserved) variables are represented by circles and manifest (observed) variables by rectangles. Single-headed arrows depict directional paths between two variables, whilst double-headed arrows depict correlations.

Box 1. Model Specification of an Unconditional Linear LGCM for a Continuous Repeated Measure

Measurement Model

$$Y_{ti} = \alpha_i + \beta_i \eta_t + \varepsilon_{ti}$$

Where $\varepsilon_{ti} \sim \mathrm{N}(0, \sigma_t^2)$

Structural Model

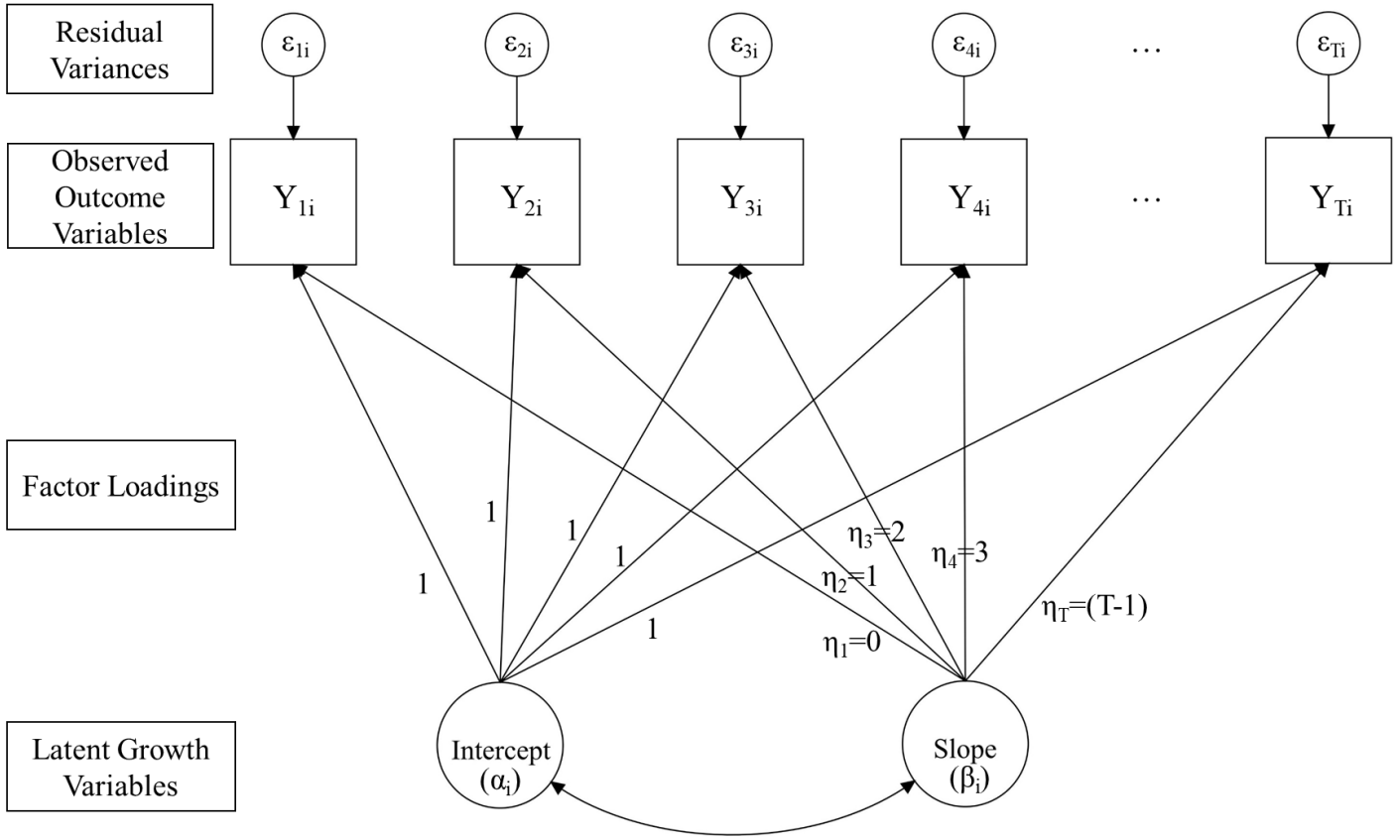
$$\alpha_i = \alpha_0 + \xi_{1i}$$
$$\beta_i = \beta_0 + \xi_{2i}$$

Where $\xi_{1i} \sim \mathrm{N}(0, \sigma_1^2)$, $\xi_{2i} \sim \mathrm{N}(0, \sigma_2^2)$,

and $\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12}^2 \\ \sigma_{12}^2 & \sigma_2^2 \end{bmatrix}$

Abbreviations: LGCM = Latent Growth Curve Model

Figure 5. Path Diagram for an Unconditional Linear LGCM for a Continuous Repeated Measure



Abbreviations: LGCM = Latent Growth Curve Model

In general, there are $t = 1$ to T time points or waves, whereby data for a continuous measure (such as cumulative days of sickness absence), Y_t, is repeatedly collected. These Y_t's are the manifest (observed) variables.

As a linear LGCM is described here, there are two latent variables: a latent intercept growth factor (α) and a latent slope growth factor (β). In the first stage of the model, known as the measurement stage, a separate linear regression line (growth curve) is fitted for each individual ($i = 1$ to N), to produce individual-specific intercepts (α_i's) and slopes (β_i's).

Furthermore, an appropriate weight, η_t, known as a factor loading in SEM terminology, is applied to the slope parameter. In the path diagram (Figure 5), an example of these factor loadings is shown where the T time points are equidistant, and have been centred so that the first time point is set to 0: $\eta_t = [0, 1, 2, 3, \dots, (T\text{-}1)]$. For example, the growth curve might relate to a study whereby repeated monthly measurements are taken up to T months. Therefore, using the specified η_t factor loadings, one unit of η_t would signify one month in this study. It is useful to centre these slope factor loadings to 0, as this facilitates interpretation of the intercept parameter, α_i (to signify the expected response at the first time point).

Slope factor loadings are to be specified by the analyst and are especially important when the $t = 1$ to T time points are not all equidistant, as the LGCM framework offers the flexibility to use different loadings, in order to enforce the linear change in the model to be proportional to these unequal intervals.

For example, suppose there are $t = 4$ time measurements in a study, taken at 1, 3, 6, and 12 months, respectively. Here it would not be appropriate to use the previously

mentioned factor loadings, which now correspond to η_t = [0, 1, 2, 3]. Rather, suitable slope factor loadings would be η_t = [0, 2, 5, 11] to reflect these unequal intervals (the first time point at 1 month is set to 0, then the distance of the remaining time points from month 1 are used as factor loadings).

The factor loadings for the intercept term, α_i, are typically all set to "1", although it is possible to amend this if required.[141]

In summary, in the measurement model, the Y_{ti} continuous measurements for each individual at *T* time points are linked, through a latent intercept (α_i) and slope (β_i). A residual error is also estimated for each individual at each time point, ε_{ti}, and it is assumed that these residuals are normally distributed with a mean of 0 and not correlated with the other residuals across time.

Finally, in the second stage of the model, known as the structural stage, the individual-specific intercepts (α_i) are pooled together to produce a mean intercept for the study population, (α_0). The latent intercept, α_i, is modelled as a random effect, whereby the random effect part, ξ_{1i}, represents the extent that the intercept for individual $\underline{i}$ (α_i) deviates from the pooled intercept (α_0).

Similarly, a study population slope (β_0) is obtained by pooling together the individual-specific slopes (β_i), and ξ_{2i} represents the extent of deviation between the slope for individual *i* and the pooled slope.

A correlation between the latent slope and intercept, σ_{12}^2, is also estimated (and is shown pictorially by the double-headed arrow in Figure 5). The distribution of individual-specific intercepts and slopes is assumed to be multivariate normal, with corresponding error terms assumed to be independent from the growth factors.

Thus, with this linear LGCM, intra-individual changes in the repeated measure Y_{ti} are captured by individual-specific intercepts (α_i) and slopes (β_i), through the measurement model. Whilst, inter-individual differences in growth curves are captured by the extent of deviation of individual intercepts (ξ_{1i}) and slopes (ξ_{2i}) relative to the study population parameter estimates, in the structural model. Both of these intra- and inter-individual differences are used to inform the pooled intercept (α_0) and slope (β_0) terms.

5.1.4 Linear LGCM with Binary Repeated Measures

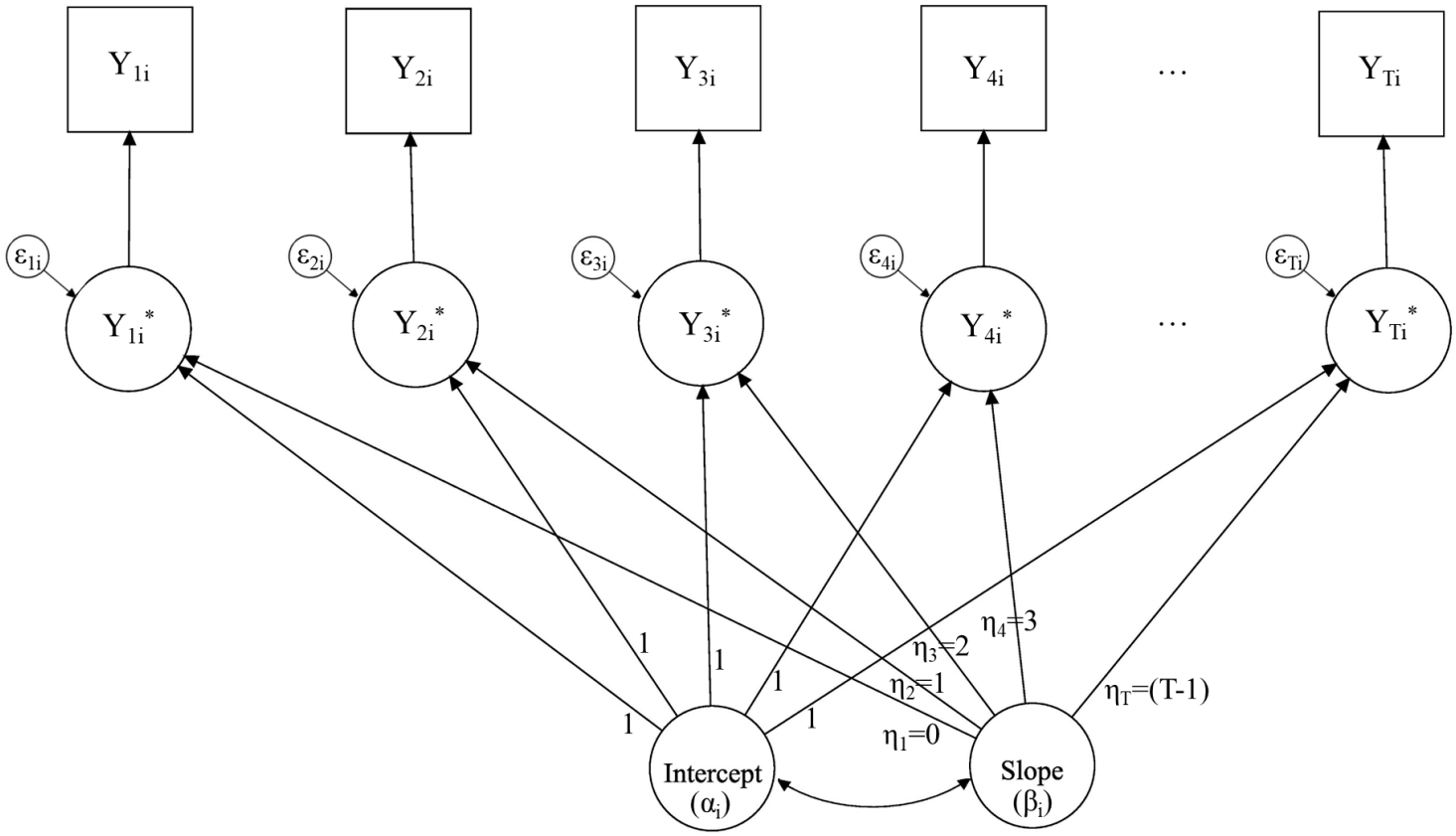
The conventional LGCM described in the previous Section uses continuous repeated measures, and the model requires these repeated measures to be normally distributed. However, some longitudinal studies may involve non-normally distributed repeated measures, such as a binary repeated measure, which often has a skewed distribution.[142]

In this thesis, as will be described later in Section 7.2.4, a binary definition of a fit note was used as the repeated measure in all trajectory models (yes/no for fit note issuance in a given time interval).

To solve the issue of distributional assumptions not being met with a binary repeated measure, a transformation can be applied prior to estimating the LGCM, to convert the binary variable into a continuous variable, whereby normality does hold. A latent response variable (LRV) transformation is apt for this purpose in a LGCM setting.[138]

A LRV transformation involves adding an extra step to the SEM path diagram for the LGCM with continuous measures from the previous Section.[138] The observed binary repeated variables (Y_{ti}) are converted into continuous latent variables (Y_{ti}*), as shown in Figure 6, and this only affects the formulae relating to the measurement model (Box 2).

Figure 6. Path Diagram for an Unconditional Linear LGCM for a Binary Repeated Measure



Abbreviations: LGCM = Latent Growth Curve Model

Box 2. Model Specification of an Unconditional Linear LGCM for a Binary Repeated Measure

Measurement Model

$$Y_{ti} = \begin{cases} 0 \text{ if } Y_{ti}^* \leq \tau \\ 1 \text{ if } Y_{ti}^* > \tau \end{cases}$$

$$Y_{ti}^* = \alpha_i + \beta_i \eta_t + \varepsilon_{ti}$$

Structural Model

$$\alpha_i = \alpha_0 + \xi_{1i}$$
$$\beta_i = \beta_0 + \xi_{2i}$$

Where $\xi_{1i} \sim N(0, \sigma_1^2)$, $\xi_{2i} \sim N(0, \sigma_2^2)$,

and $\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12}^2 \\ \sigma_{12}^2 & \sigma_2^2 \end{bmatrix}$

Abbreviations: LGCM = Latent Growth Curve Model

To perform this LRV transformation, it is assumed that the observed Y_{ti} values are a binary form of an underlying continuous LRV, Y_{ti}*. The relationship between the binary Y_{ti} and continuous Y_{ti}* values is defined by a cut-point (or threshold), τ, whereby Probability(Y_{ti} = 1) = Probability(Y_{ti}* > τ), as shown in Box 2. In the context of this thesis, the LRV transformation is such that the probability of a fit note being issued for individual *i* during time interval *t*, is equal to the probability that the LRV Y_{ti}* is greater than this threshold τ. The threshold, τ, is generally defined to be a constant and therefore invariant over time (known as the threshold invariance assumption).[138,143]

The measurement model (Box 2) then uses these transformed latent Y_{ti}* values (instead of observed Y_{ti} values) in the same manner as in the previous Section (the same notation has been used too), except that the distribution of the error term, ε_{ti}, now needs to be defined by the user to reflect this LRV transformation. The two most common

distributions are a standard logistic distribution (this is known as a logit link function) or a standard normal distribution (a probit link function).[138] Logistic regression results using a logit or probit link function are often indistinguishable.[144]

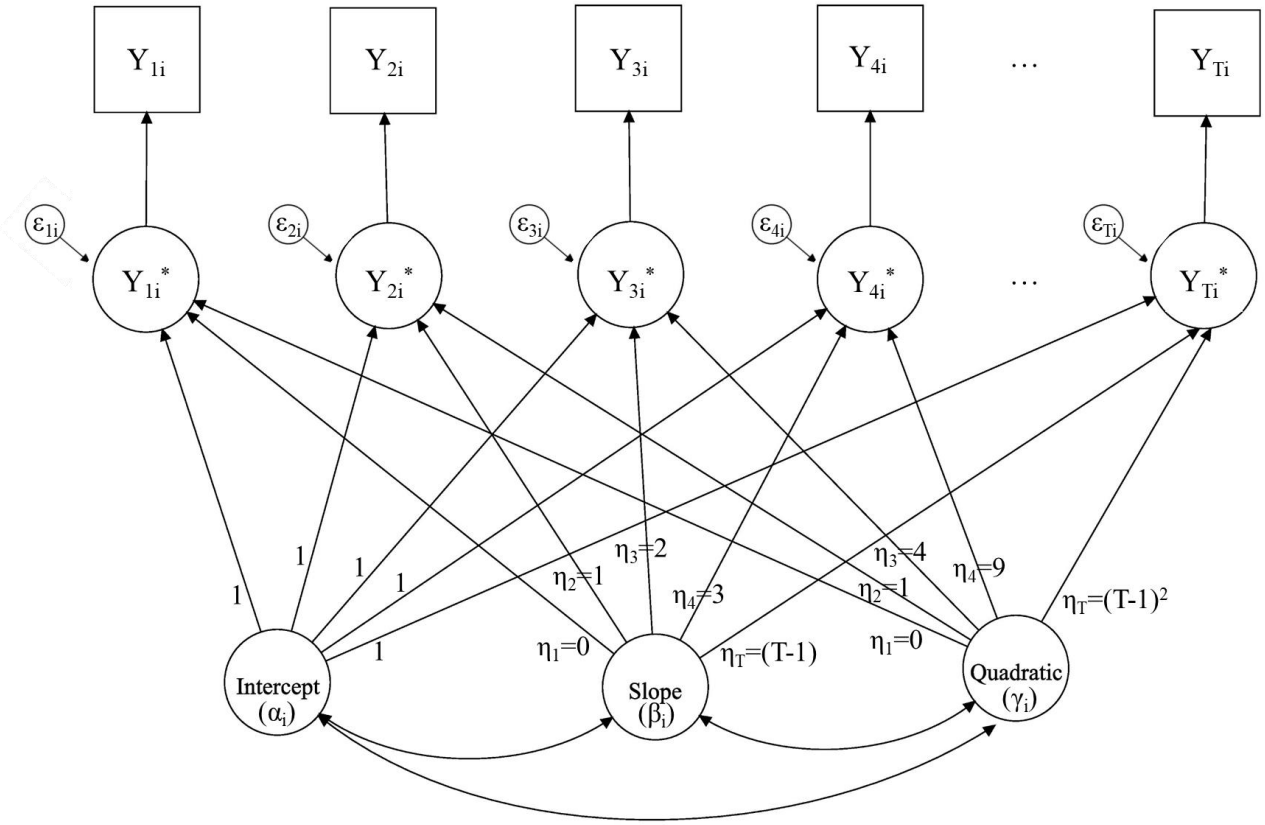
Finally, the remainder of the LGCM process (i.e., the structural model) is the same as Section 5.1.3.

5.1.5 LGCM Extensions

Thus far the most basic form of a LGCM has been considered: a linear LGCM, whereby the development of the repeated measure over time was considered to occur in a linear fashion. However, extensions exist to allow for more complicated patterns of development over time, such as higher order polynomials.

For example, there may be theoretical reasons to suggest that the growth curves of the repeated measure are better represented using a quadratic shape. In this case, a quadratic LGCM based on binary measures extends the equivalent linear model in the previous Section, by adding a third latent growth parameter for the quadratic term (as shown in the path diagram in Figure 7 and with accompanying equations in Box 3). The latent quadratic growth term is represented as γ_i in the measurement model, and has factor loadings (η_t) that are squared. Furthermore, in the structural model an average quadratic term, γ_0, is estimated for the study population, with inter-individual variance estimated as ξ_{3i}.

Figure 7. Path Diagram for an Unconditional Quadratic LGCM for a Binary Repeated Measure



Abbreviations: LGCM = Latent Growth Curve Model

Box 3. Model Specification of an Unconditional Quadratic LGCM for a Binary Repeated Measure

Measurement Model

$$Y_{ti} = \begin{cases} 0 \text{ if } Y_{ti}^* \leq \tau \\ 1 \text{ if } Y_{ti}^* > \tau \end{cases}$$

$$Y_{ti}^* = \alpha_i + \beta_i \eta_t + \gamma_i (\eta_t)^2 + \varepsilon_{ti}$$

Structural Model

$$\alpha_i = \alpha_0 + \xi_{1i}$$
$$\beta_i = \beta_0 + \xi_{2i}$$
$$\gamma_i = \gamma_0 + \xi_{3i}$$

Where $\xi_{1i} \sim N(0, \sigma_1^2)$, $\xi_{2i} \sim N(0, \sigma_2^2)$, $\xi_{3i} \sim N(0, \sigma_3^2)$,

and $\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12}^2 & \sigma_{13}^2 \\ \sigma_{12}^2 & \sigma_2^2 & \sigma_{23}^2 \\ \sigma_{13}^2 & \sigma_{23}^2 & \sigma_3^2 \end{bmatrix}$

Abbreviations: LGCM = Latent Growth Curve Model

The functional form of a LGCM can be further extended, to cubic or beyond if desired. However, it must be noted that a minimum of three time points are needed to model a linear LGCM, four time points for a quadratic LGCM, five for a cubic LGCM, etc.[145] Hence choice of the polynomial functional form of the LGCM may be influenced by the number of repeated time measurements available.

Yet another alternative LGCM specification available for use is a piecewise growth model (also known as a slope segment model).[141] A piecewise model is especially useful if different rates of change in the repeated measure are expected between different time points.

For example, in a study comparing short and long-term effects following a drug intervention, a strong positive rate of change may be expected in the first 3 months of follow-up, and a weaker positive or even negative rate of change from months three to twelve. A piecewise model also offers the flexibility to choose how many "pieces" to

break the time continuum into, and the analyst can specify combinations of different polynomial functional forms for different "pieces" if required. Factor loadings of the different slope parameters for each "piece" need to be specified appropriately.

An example piecewise LGCM based on repeated binary measures and two linear slopes is shown in Figure 8. This piecewise model was used in Chapter 7 (mentioned again in Sections 7.3.4 and 7.4.4), whereby a LGCM based on a binary definition of fit note issuance was modelled with a linear slope for the first 12 months of follow-up (β_{1i}), and a different linear slope for years two to three of follow-up (β_{2i}). This allowed separate slopes to be estimated for shorter and longer-term absences.

Factor loadings were chosen such that each time unit represented one month. There were four three-monthly recurring time intervals in the first slope-segment, and four six-monthly recurring time intervals in the second slope-segment (these interval lengths guided the selection of factor loadings in Figure 8).

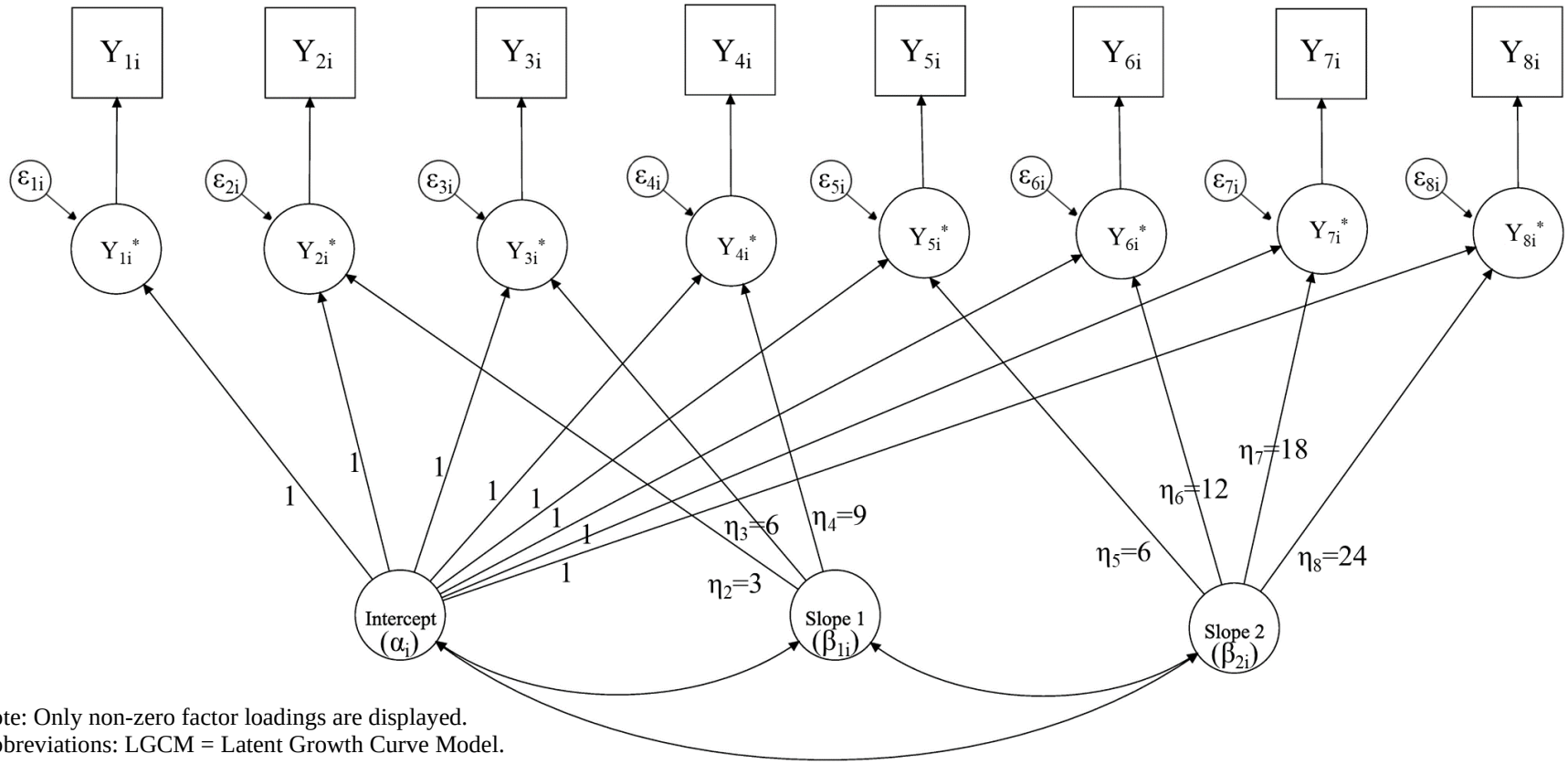
Finally, additional LGCM extensions include:

- Estimating separate growth curves for different subgroups of the study population (this is important for this thesis, and detailed in Section 5.2),

- Using a conditional model to investigate the effect of a predictor variable(s) on the growth curve (this is also addressed in Section 5.2), and

- Basing the LGCM on other non-normally distributed variables, such as a count variable

Figure 8. Path Diagram for an Unconditional Linear Piecewise LGCM for a Binary Repeated Measure



Note: Only non-zero factor loadings are displayed.
Abbreviations: LGCM = Latent Growth Curve Model.

5.1.6 Model Fit Evaluation

An important question during the LGCM modelling process, is to ask how well the model fits the data. A variety of LGCM specific model fit indices exist in this regard. In this thesis, several of the most important model fit indices as recommended by Wickrama et al (2021) were utilised:[141]

- The Root Mean Square Error of Approximation (RMSEA);[146] values closer to 0 indicate better model fit
- Comparative Fit Index (CFI);[147] ranges from 0-1, higher values indicate better fit
- Tucker-Lewis Index (TLI);[148] ranges from 0-1, higher values indicate better fit
- Standardised Root Mean Residual (SRMR);[149] ranges from 0-1, lower values indicate better fit

Specifically, these model fit indices were used alongside cut-off guidelines from a simulation study performed by Hu and Bentler (Table 8),[150] to assess the quality of all LGCMs fitted in this thesis.

Furthermore, closeness of model fit to the data was also evaluated graphically, through inspection of the model estimates (in the context of this thesis, this relates to probabilities of fit note issuance at each time interval, as explained further in Chapter 7), compared to the observed data (proportions of fit note issuance).

Table 8. Criteria to Assess Quality of LGCM Model Fit in this Thesis

Model Fit Index	Cut-Off
RMSEA	≤ 0.08
CFI	≥ 0.95
TLI	≥ 0.95
SRMR	≤ 0.06
Graphical Inspection	Estimated probabilities compared against observed fit note issuance proportions (less visual difference suggests better fitting model)

Abbreviations: LGCM = Latent Growth Curve Model; RMSEA = Root Mean Square Error of Approximation; CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; SRMR = Standardised Root Mean Residual

5.1.7 Main Limitation of LGCM

The main limitation of a LGCM model is that all individuals within a population are assumed to follow a similar development of the repeated measure over time (i.e., a one-size-fits-all trajectory is assumed). If this assumption is reasonable given the data being analysed, a LGCM can perform well in capturing the interindividual variability in the repeated measure.

However, this assumption might not hold, and there might be underlying heterogeneity in the trajectories of individual study participants. If heterogeneity is present, it may be observed, through statistically significant variance growth parameters estimated in the LGCM, or it may be concealed. For example, there could be two distinct subgroups of individuals in a study, one subgroup that has a trajectory of increasing growth, and the other a decreasing growth rate. When assessing the overall study LGCM, one-size-fits-all trajectory, this may indeed appear stable over time (i.e., indicating no growth), as the two distinct opposing trajectories may cancel each other out.

If heterogeneity does exist in the LGCM (either unobserved or observed), a better approach would be to consider estimating multiple trajectories rather than a single

average trajectory, whereby subgroups of individuals are assigned to the study-level trajectory that their own individual trajectory is most similar to. Two extensions of a LGCM, LCGA and GMM, are designed for this purpose and described in the next Section.

5.2 Finite Mixture Models

LCGA and GMM are known as finite mixture models, such models are adept at analysing repeated measures in situations involving a finite number of homogenous subpopulations.[151,152] LCGA and GMM models were developed by Muthén and Shedden in 1999, whereby the LGCM was extended by having finite mixture modelling applied to it, which allows two or more LGCMs to be used in predicting subgroup variability in trajectories.[153] The basic principle, as indicated by Muthén, is to consider "each GCM as modelling a separate subpopulation following a different growth curve"[85(p115)], thus for each subgroup (or class or cluster) of the population, a distinct average growth curve is estimated.

A LCGA model is a simplified version of a GMM, and is described in the next Section. For simplicity, models in this Section are presented based on a linear polynomial functional form and with binary repeated measures (as this is the type of repeated measure used in this thesis).

5.2.1 Latent Class Growth Analysis (LCGA)

A LCGA model (also referred to as 'group-based trajectory modelling'),[62] is semi-parametric, whereby separate, homogenous subgroups of individuals with similar patterns of repeated measure behaviour over time (hereafter referred to as trajectory classes) are estimated for a study population. The probability of belonging to a

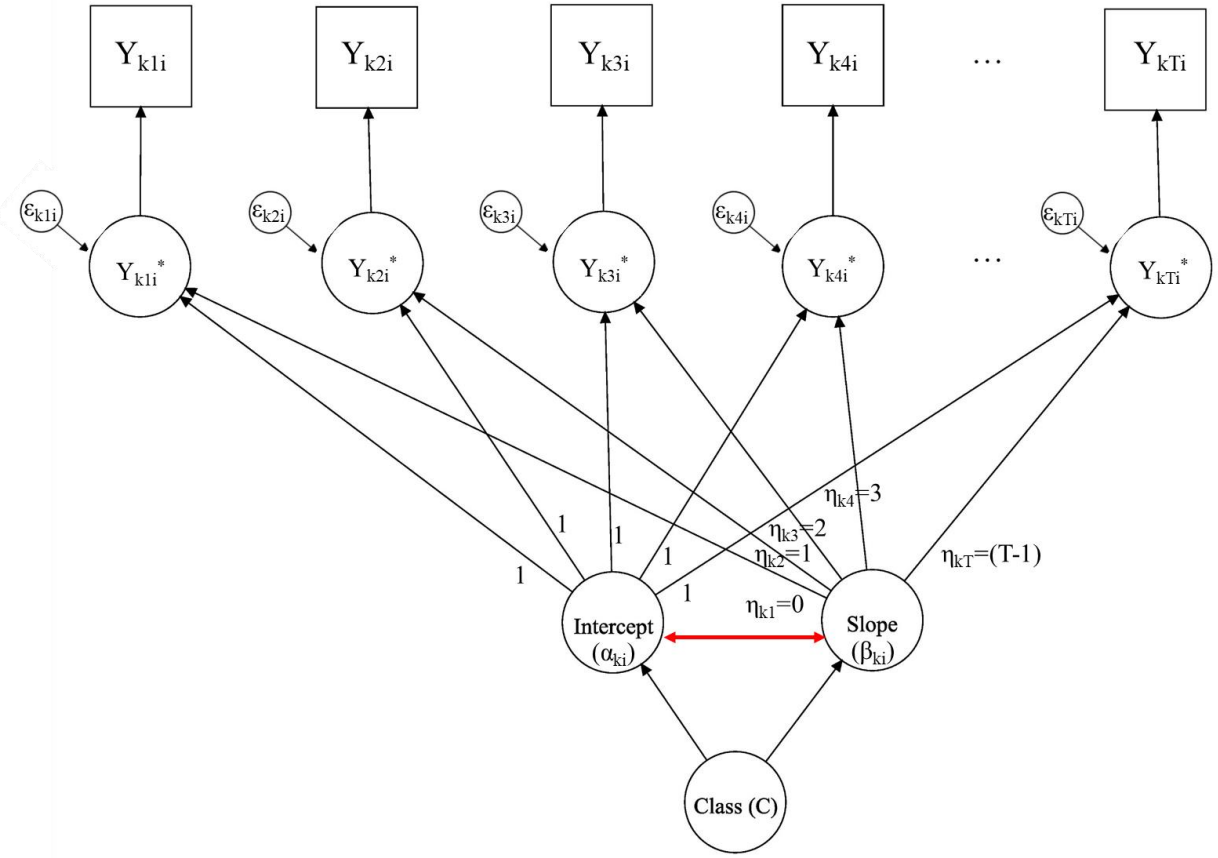
trajectory class (or cluster), known as posterior probability, is estimated for each individual using the observed data, and the individual is assigned to the class that they have the greatest probability of belonging to.

Separate mean intercepts and slopes are estimated for each trajectory class, as shown in Figure 9 and Box 4. The notation is the same as for the linear LGCM for binary measures (Section 5.1.4), except that all the parameters are now indexed with a '*k*', as separate parameters are now estimated for each of $k = 1$ to C trajectory classes. The trajectory class variable, *k*, is categorical and latent (i.e., unobserved). Hence in a LCGA, the latent variables estimated are now a combination of continuous (i.e., the growth intercept and slope parameters) and categorical variables (the trajectory classes).

Furthermore, in a LCGA, whilst the measurement model is the same as the corresponding LGCM in Section 5.1.4 (with the exception of being partitioned by trajectory class), in the structural model, homogeneity of classes is enforced (and terms are also indexed by class). That is, unlike in a LGCM, no inter-individual variation is permitted within a (class specific) GCM here, and a fixed effect is applied when pooling the individual (and class)-specific intercept and slope terms (to produce α_{k0} and β_{k0}, respectively). In other words, all individuals are assumed to follow their estimated trajectory class.

Therefore, it follows that in the path diagram, there is now no correlation between the latent intercept and slope terms either (Figure 9).

Figure 9. Path Diagram for an Unconditional Linear LCGA/GMM for a Binary Repeated Measure



Abbreviations: LCGA = Latent Class Growth Analysis; GMM = Growth Mixture Model.
Note: A correlation between the latent intercept and slope parameters (red double-headed arrow) is only permitted for a GMM.

Box 4. Model Specification of an Unconditional Linear LCGA/GMM for a Binary Repeated Measure

Measurement Model

$$Y_{kti} = \begin{cases} 0 \text{ if } Y^*_{kti} \leq \tau \\ 1 \text{ if } Y^*_{kti} > \tau \end{cases}$$

$$Y^*_{kti} = \alpha_{ki} + \beta_{ki}\eta_{kt} + \varepsilon_{kti}$$

Structural Model

$$\alpha_{ki} = \alpha_{k0} + \xi_{k1i}$$
$$\beta_{ki} = \beta_{k0} + \xi_{k2i}$$

Where $\xi_{k1i} \sim N(0, \sigma^2_{k1})$, $\xi_{k2i} \sim N(0, \sigma^2_{k2})$,

and $\boldsymbol{\Sigma}_k = \begin{bmatrix} \sigma^2_{k1} & \sigma^2_{k12} \\ \sigma^2_{k12} & \sigma^2_{k2} \end{bmatrix}$

A LCGA model is a special case of a GMM, whereby the parameters in red text are set to equal 0 (i.e., the within class variances of the growth parameters are removed from the structural model and only means are estimated):

$$\alpha_{ki} = \alpha_{k0}$$
$$\beta_{ki} = \beta_{k0}$$

Abbreviations: LCGA = Latent Class Growth Analysis; GMM = Growth Mixture Model.
Note: Within class variances of the latent intercept and slope parameters (red text) are only permitted for a GMM.

5.2.2 Growth Mixture Modelling (GMM)

A GMM is fully parametric, and here the assumption of homogenous trajectory classes is relaxed. Unlike with a LCGA, within-class heterogeneity is now permitted with a GMM, through use of random effects.[154,155] That is, some variation is now permitted around the class-specific trajectory for individuals within a class, rather than assuming all individuals follow a common class-specific trajectory.

The GMM measurement and structural model is identical to the LCGA, except that class-specific random effects of inter-individual variation in growth curves are now

estimated for the latent intercepts (ξ_{k1i}) and latent slopes (ξ_{k2i}), as shown in Figure 9 (now including the red text). Correlation between these latent growth parameters is also permitted, as shown in Box 4 (red double-headed arrow).

There are two approaches to estimating these random effects in a GMM. Firstly, the simpler approach is a GMM Class Invariant (GMM-CI) model. Here the variances of the trajectory slope and intercept parameters are relaxed to take non-zero values, but constrained to take the same value, independent of trajectory class. That is, the variance of individual intercept terms is fixed across classes ($\sigma_{k1}^2 = \sigma_1^2$ for all classes, k = 1 to C), as is the variance of individual slope terms ($\sigma_{k2}^2 = \sigma_2^2$ for all classes, k = 1 to C). However, simulation studies have shown that parameter estimates can be biased when using a GMM-CI approach.[156–158]

The second option, a GMM Class Variant (GMM-CV) model, is more complex, and allows the variance of the trajectory slope and intercept parameters to be non-zero and free across trajectory classes. That is, the intercept and slope variance terms remain unchanged as in Box 4, specific to each class k (as σ_{k1}^2 and σ_{k2}^2, respectively).

Furthermore, a decision can be made whether to free the variances for all k = 1 to C classes, or to use a combined approach, with a GMM-CV that has some classes with freed variance, and others with either a common variance or even a variance of zero. This flexibility of the GMM-CV also extends to modifying the covariances of these parameters.

5.2.3 Comparing LCGA and GMM Models

Both the LCGA and GMM are indeed similar in that they estimate distinct trajectories within a study population. A LCGA is a nested version of a GMM in its most reduced

form (a GMM-CI model is then more complex, and a GMM-CV the most complex). The only difference between a LCGA and GMM concerns assumptions around individual variability within trajectory classes.

Within-class variability is permitted in a GMM (through random effects of class-specific latent growth parameters), but not with a LCGA. Therefore, if within-class variability of individuals is close to zero for all classes, the GMM and LCGA models will be similar. However, if this is not the case, the GMM offers a potentially more realistic interpretation of how individuals actually behave within their assigned subgroup and is especially important if there are few subgroups.

To illustrate this difference in assumptions, consider a hypothetical study with three distinct RTW trajectory classes of: a rapid RTW, a slow RTW, and sustained work absence. Here, with a LCGA, individuals would be fixed to these three trajectory classes, and individual variation within the same trajectory class would not be permitted.

Whereas application of GMM offers more flexibility and does permit individual within-class variability. For example, individuals in the rapid RTW trajectory would in general be characterised by a similar fast RTW (at least, faster than the rest of the study population), but some individuals in this trajectory could have a very fast RTW, and others a less fast RTW.

Nonetheless, whilst the GMM may offer more realistic derived trajectory classes, an important difficulty of such models is the increased complexity in constructing and later interpreting them. This increased complexity also results in greater computational burden, which can often cause model convergence issues.[62] There are also more

decisions regarding model choices to be made with a GMM, such as choosing whether to free the latent parameter variances to non-zero values:

- separately for all classes (a GMM-CV model),
- separately for certain classes (a reduced version of a GMM-CV model),
- or as a common value across classes (GMM-CI model).

On the contrary, the LCGA is the simpler model. There are less parameters to be estimated, which reduces computational burden and increases the likelihood of model convergence.[159] Furthermore, the increased simplicity of the model can make the results easier to interpret, especially for a non-technical audience. The LCGA is often deemed to be a more practical model choice than GMM for researchers, and is widely used.[62,160] For example, in the six studies identified in this thesis' systematic review of trajectories of work absence, three studies used LCGA and none used GMM (a narrative synthesis of trajectory derivation methodology from these studies is described later in Section 5.4).

In this thesis, considering the overall merits and limitations of LCGA and GMM models, both were used when deriving trajectories of work absence. The specific model building strategy used, considers LGCM, LCGA and GMM, and is explained later in Section 7.2.5.

5.2.4 Model Convergence

There can often be model convergence issues when estimating mixture models. These issues are generally exacerbated when there are a greater number of trajectory classes in the model, and when the model itself is more complex. For example, as already

mentioned, more convergence issues may be anticipated if a GMM model is used, rather than a LCGA model.[141]

One common convergence issue is when the log-likelihood, which is maximised during the model fit process, converges to a "local maxima" rather than the true global solution.[161] This issue can be overcome by increasing the number of starting values and iterations in the process of maximising the log-likelihood.[141] Jung and Wickrama proposed that between 100 to 500 random sets of starting values are used to achieve the true global maximum log-likelihood.[61] Therefore, in an effort to limit convergence issues, for all LCGA and GMM models fitted in this thesis, 500 random sets of starting values were used, with 100 iterations.

Furthermore, for models that converged, a secondary analysis was also performed in this thesis to affirm replication of model results, as is recommended.[141] This replication involved fitting the same model twice more, using seed values from the top two highest log-likelihood values. If model results remained unchanged across the three models, then it could be concluded that there was no evidence of a "local maxima" solution.

5.2.5 Model Fit Evaluation

When assessing performance of competing LCGA and/or GMM models, the model fit indices previously described for a LGCM (Table 8) are not applicable (RMSEA, CFI, TLI and SRMR). These model fit indices are based on chi-square values, but as the number of classes in LCGA and GMM models is greater than 1, the model data cannot be fitted to a single covariance matrix, which prevents these indices from being estimated.[141,162]

Instead, other indices that are relevant for mixture modelling were used in this thesis.

Firstly, there were two Information Criteria (IC) statistics:

- The Akaike Information Criterion (AIC);[163] lower values indicate better model fit
- The Bayesian Information Criterion (BIC);[164] lower values indicate better model fit

Other indices for comparing model fit include types of Likelihood Ratio Test (LRT). The two LRTs commonly used to aid decision-making regarding the optimal number of trajectory classes,[141] are:

- The Lo-Mendell-Rubin LRT (LMR-LRT)[165]
- The bootstrapped LRT (BLRT)[151]

Both the LMR-LRT and BLRT were used in this thesis. These tests compare the current (*k* class) LCGA model, against a corresponding LCGA model with one fewer class (i.e., *k*-1 classes), and are based on differences in log-likelihood values. A *p* value is then obtained from these LRTs, and if statistically significant (< 0.05), is indicative that the current (*k* class) model fits the data better than the *k*-1 class model.

Unlike the LMR-LRT, the BLRT is based on a re-sampling procedure, known as bootstrapping. In this thesis, 100 bootstrap draws were used when calculating the BLRT, as is recommended in order to generate a more accurate *p* value from this type of LRT.[151] Generally, as shown through simulation, the BLRT performs better than the LMR-LRT and types of IC statistic as a model fit index.[166,167] Albeit the BLRT tends to increase model computational time, and is sensitive to more complex models.

Aside from these four model fit indices, model fit of LCGA and GMM models was also assessed graphically. The same approach was used as for LGCM model fit evaluation,

whereby plots of model estimated probabilities were contrasted against observed fit note issuance proportions.

The remaining criteria used to evaluate competing models mainly concerned the suitability of the trajectory classes derived, as described in the next Section. The final set of guidelines used to inform choice of an optimal LCGA or GMM in this thesis are summarised in Table 9.

Table 9. Guidelines Used for Assessing Optimal Class LCGA/GMM Model

	Measure	Criteria
Model Fit	AIC	Lower values
	BIC	Lower values
	LMR-LRT	$p < 0.05$
	BLRT	$p < 0.05$
	Graphical Inspection	Estimated probabilities compared against observed fit note issuance proportions (less visual difference suggests better fitting model)
Meaningfulness of Classes	Average Posterior Probabilities of Classes	≥ 0.7
	Entropy	≥ 0.7
	Class Prevalence	$\geq 1\%$
	Graphical Inspection	Class shapes evaluated for interpretability, separability, and face validity
Parsimony		Model with k classes chosen over (k+1) class model if above metrics were similar

Abbreviations: LCGA = Latent Class Growth Analysis; GMM = Growth Mixture Model; AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; LMR-LRT = Lo-Mendell-Rubin Likelihood Ratio Test; BLRT = Bootstrapped Likelihood Ratio Test.

5.2.6 Assessing Meaningfulness of Classes

To assess the degree of differentiability of derived trajectory classes in all LCGA and GMM models in this thesis, two different indices were used: average posterior probability, and entropy. These indices are now described.

When a LCGA or GMM model is fitted, each individual is assigned a posterior probability for each trajectory class, which represents their likelihood of belonging to that particular class. To illustrate this, an example is shown in Table 10, based on a fictitious dataset of n=100 individuals. Here a five-class LCGA model was fitted and during this process each individual received a posterior probability for each trajectory class. Then, each individual was assigned to the trajectory class that they had the greatest probability of belonging to.

For example, the participant with ID = 1, had a high probability of belonging to class five (posterior probability = 0.87), and low probability of belonging to any other class (≤ 0.10 in all cases). This participant was ultimately assigned to class five (as they had the highest posterior probability for this class).

Table 10. Fictitious Example of Posterior Probabilities Based on a Five-Class LCGA Model

	Posterior Probability					
Participant ID	**Class 1**	**Class 2**	**Class 3**	**Class 4**	**Class 5**	**Final Assigned Class**
1	0.00	0.02	0.10	0.01	**0.87**	5
2	0.00	**0.77**	0.10	0.13	0.00	2
3	**0.55**	0.00	0.45	0.00	0.00	1
4	**0.85**	0.00	0.05	0.00	0.10	1
5	0.12	0.00	**0.83**	0.00	0.05	3
6	0.01	0.15	0.00	**0.77**	0.07	4
7	**0.95**	0.00	0.00	0.03	0.02	1
8	0.01	**0.85**	0.11	0.01	0.02	2
9	0.02	0.00	0.04	0.00	**0.94**	5
…	…	…	…	…	…	…
100	0.00	0.02	0.08	**0.90**	0.00	4

Based on a fictitious dataset of n=100 individuals.
The highest posterior probability for each individual is shown in bold text.

An example where the trajectory class assignment was less clear, was shown for participant ID = 3. Here there was a posterior probability of 0.55 of belonging to class

one, and 0.45 for belonging to class three. Ultimately, the participant was assigned to class one, due to the (slightly) higher posterior probability, albeit these two classes were not as clearly separable for the class allocation of this participant.

A trajectory class that is clearly distinguishable from remaining classes is characterised by containing more individuals with high posterior probabilities for that particular trajectory class (closer to 1) and low posterior probabilities (closer to 0) for the other classes. The extent of trajectory class separation can be formally evaluated through taking the average posterior probability for each trajectory class (i.e., by taking an average of the bold values for each class posterior probability column in Table 10). For this fictional dataset, the average posterior probabilities are shown in Table 11.

Table 11. Average Posterior Probabilities for Fictitious Example Based on five-Class LCGA Model

Average Posterior Probability				
Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
0.78	0.81	0.83	0.84	0.91

Entropy is then calculated as a standardised measure ranging from 0 to 1, based on these average posterior probabilities.[168] An entropy of 0 indicates that there is no class separability in the LCGA or GMM (i.e., for any given individual, they have equal chance of belonging to any of the derived classes), whilst an entropy of 1 indicates perfect class separability (i.e., individuals have a posterior probability of 1 for the class they are assigned to and 0 for remaining classes).

Although higher values of entropy or average posterior probabilities are indicative of better class separability, there are currently no formally agreed cut-off criteria for evaluating this in the literature.[61] Nonetheless, as a guide for decision-making when comparing competing LCGA and GMM models in this thesis, entropy and average

posterior probabilities ≥ 0.7 were considered to demonstrate good class separability in this study, in line with the suggestion of Andruff et al.[169]

Another important feature of LCGA and GMM models to assess, is the proportion of the sample that is contained within each class, i.e., trajectory prevalence. When a LCGA or GMM model is fitted, trajectory classes are derived based on posterior probability. As shown in Table 12 (based on the previously mentioned fictitious example), a trajectory class count based on posterior probability (i.e., this is derived from summing the posterior probability columns for each class in Table 10), can be slightly different to a class count based on most likely latent class membership.

If entropy is 1 and the classes are perfectly separable, then there will be no difference between the two count methods. In this thesis, unless otherwise stated, trajectory prevalences are reported based on posterior probabilities, as this is what the classes are constructed from.

Table 12. Trajectory Prevalence Based on Posterior Probability Counts Compared to Most Likely Latent Class Membership for Fictitious Example

Count Based On:	**Class 1**	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Posterior Probability	25.1 (25.1%)	18.1 (18.1%)	17.6 (17.6%)	18.5 (18.5%)	20.7 (20.7%)
Most Likely Latent Class Membership	30 (30.0%)	20 (20.0%)	10 (10.0%)	20 (20.0%)	20 (20.0%)

N (%) is shown, where N is the number of individuals belonging to a particular trajectory class, and the % is then calculated from the total of n=100 individuals

It has been suggested that an additional prerequisite for a well performing model should be a trajectory prevalence of ≥ 5% for all classes (or alternately, it is recommended that there are ≥ 25 individuals in any given trajectory class).[170,171] Where trajectory classes are included with a smaller prevalence than this, the researcher should be able to defend their inclusion based on the meaningfulness of the trajectory classes. These

recommendations originate from concerns that these smaller trajectory classes might arise due to a low powered trajectory derivation analysis, and therefore possibly being spurious classes that might not be replicated in other similar datasets.

Originally, it was planned to also apply the same suggested criteria in this thesis, of a minimum requirement of $\geq 5\%$ trajectory prevalence when deciding on an optimal model. However, after observing the trajectory shapes of the smallest classes derived in this study, and considering the high power of trajectory prevalence estimation demonstrated later in Section 7.2.2, this was relaxed to a requirement of $\geq 1\%$ trajectory prevalence per class. The CPRD data request protocol was thus amended to reflect this change also (as mentioned in Section 4.3).

Additionally, the different trajectory shapes identified in this thesis were visually evaluated for interpretability, separability (for example, visual presence of similar shaped trajectory classes was explored), and meaningfulness in relation to possible expected sickness absence patterns from the literature and feedback from a HCP, the Office for Health Improvement and Disparities (OHID, formerly known was Public Health England), and a patient and public involvement and engagement (PPIE) group.

Finally, when comparing two competing LCGA or GMM models with k and k+1 classes, if all of the criteria mentioned thus far are similar, it is recommended that the more parsimonious model (i.e., with fewer trajectory classes) is usually elected, unless the researcher can otherwise defend inclusion of the higher class model.[172]

Therefore, in this Section and the previous Section, it has been demonstrated that choosing an optimal LCGA or GMM model requires a holistic approach. Consideration must be given towards statistical measures of model fit, as well as the interpretability

and separability of the derived classes, and decisions must be overlaid with theoretical expectations.

5.2.7 Further Considerations When Performing LCGA or GMM

If the performance of a linear LCGA or GMM is not deemed sufficient, similar to LGCM, LCGA and GMM models can also be extended by considering polynomial functional forms beyond linear, such as quadratic or cubic functional forms. Furthermore, approaches such as piecewise modelling can also be applied to LCGA or GMM (see analogous application for LGCM in Section 5.1.5).

Another model extension, which is the main focus of Chapter 8, is a conditional LCGA or GMM, whereby covariates are added to the model. This allows for the presence of any associations between covariates and trajectory membership to be tested. This type of analysis is explained in more detail in Section 8.3.

More generally, when performing LGCM, LCGA or GMM analysis, there are also a range of different statistical software that can be used, such as Mplus, Stata, and SAS, to name a few.

In this thesis, Mplus (Version 8.9) was used for all trajectory derivation and trajectory-covariate association analysis. Mplus is widely used for trajectory derivation research, for example, in a 2016 systematic review of LCGA and GMM trajectory studies,[173] n=29 studies (81%) used Mplus. Furthermore, all of the course materials from the main LCA training course that AL attended during this thesis were Mplus-based, hence this choice was also for reasons of practicality.

Mplus is also advantageous in instances of incomplete follow-up data, as it allows use of an estimation approach called Full Information Maximum Likelihood (FIML) to

incorporate the non-missing repeated measurement data of such individuals (i.e., individuals are retained in the analysis as long as they have some non-missing data).[141] Therefore this approach can increase the number of individuals that can contribute to a study by, for example, including those who drop-out of the study before end of follow-up.

5.3 Other Individual-Centred Longitudinal Methods For Consideration

Two further individual-centred approaches for analysing longitudinal repeated measures data are now described for completeness. These are latent transition analysis (LTA) and sequence analysis. Both of these trajectory derivation methods were used in the included studies of the systematic review in Chapter 3 and were considered for this thesis, although ultimately not used (this will be explained further in Section 5.5).

5.3.1 Latent Transition Analysis (LTA)

LTA is a type of finite mixture model (akin to LCGA and GMM), and is semi-parametric (like LCGA).[174] However, the true novelty of LTA, and difference from LCGA and GMM, is that it permits individuals to change from one identified trajectory class to another over the course of time (i.e., 'transition' is allowed). Indeed, LTA models have similarities with Hidden Markov Models, which are special statistical approaches commonly used in biological sequence modelling.[62,175]

As an example, using LTA would allow an individual who is in a trajectory class of continuous sustained absence at time point *t*, to transit into a different trajectory class characterised by sporadic absence at the next time point *t*+1. Further still, at time point time point *t*+2, the same individual could also transit back into the sustained absence trajectory class.

The main emphasis of LTA is to study these changes between different trajectory classes across time. However, LTA is a complex and computationally demanding model. In particular, change between classes across two consecutive time-points is estimated through a matrix of transition probabilities (this is unique to LTA), and these matrices (one for every pair of consecutive time points) require estimation of a large number of parameters.[62] Therefore, a limitation of LTA is that a large sample size is often needed for the model to run efficiently and to maintain sufficient power.[62] In addition, when there are more follow-up time points, these models increase considerably in complexity.[62]

As LTA originates from the same family of models as LCGA and GMMs and is also a type of LCA, choice of an optimal class LTA model can be guided using the same model convergence and fit statistics mentioned previously for LCGA and GMM (see Sections 5.2.4 and 5.2.5). Furthermore, missing data can also be handled using the same approach as for LCGA and GMM models (Section 5.2.7).

5.3.2 Sequence Analysis

Another individual-centred approach to longitudinal data modelling, is sequence analysis. However, in contrast to the types of LCA discussed thus far in this Chapter, sequence analysis is fully non-parametric and does not make any distributional assumptions.[62] Unlike with the finite mixture models mentioned thus far that base trajectory allocation on conditional probabilities of trajectory membership, sequence analysis groups individuals into trajectories based on similarity of their sequences (ordered patterns of discrete states over time).[62] Furthermore, sequence analysis can only be performed on categorically defined repeated measures (not continuous definitions).

As an example of possible types of sequences, consider a hypothetical study with data collected monthly, over a one-year follow-up, and two possible states, S (for sickness absence) and R (for RTW). A sequence of SSSSSSSSSRRR would represent that an individual spent the first nine months of follow-up in a sustained sickness absence state, followed by a sustained RTW for the final three months (this could be termed a 'late RTW' trajectory). Another individual may have had a different sequence, still containing nine months of absence and three months of RTW: SSSRRRSSSSSS, this would represent a RTW after three months, followed by a relapse at the start of month seven (this could be termed 'early relapse').

When performing sequence analysis, there are typically two steps, the first concerns using a distance measure to compare the similarity of all the different observed sequences in the study population, then in the second step, this distance measure is used in a process to group individuals into homogenous subgroups based on similarity of their sequences.

The most common type of distance measure used is optimal matching; indeed, the terms 'sequence analysis' and 'optimal matching' are often used interchangeably.[62] Optimal matching uses an 'update-based' approach, whereby an algorithm computes the minimum number of operations (defined as addition, subtraction, or substitution of states within a sequence) that are needed to transform the sequence from one individual, into the sequence of another.[86]

However, it must be noted that during application of optimal matching, specification of relative weights or 'costs' of operations is required.[176,177] To provide some context, referring back to the example above of 'late RTW' and 'early relapse' sequences, the minimum number of operations required to complete the transformation of these

sequences to a third sequence of sustained sickness absence (SSSSSSSSSSS) would be three substitutions in both cases (with the three R's substituted for S's). Thus, if equal weights were used for these substitutions, irrespective of where in the sequence these R's occurred, then these two sequences representing a late RTW and early relapse would be treated the same, in terms of relative 'distance' from a sustained absence trajectory. This can lead to such sequences being ultimately grouped into the same common trajectory, yet the late RTW and early relapse likely represent qualitatively different subgroups of sickness absent patients. To remedy this, the analyst can specify costs, a priori, to assign to operations in the optimal matching process. These costs can relate to the order of states within a sequence (applicable to the example provided), as well as the number and duration of states.

However, one of the main limitations of sequence analysis is that choice of these relative costs that are assigned to operations is largely arbitrary, and this selection can highly affect the trajectory derivation results.[176,177] Thus results may not be generalisable across studies if authors use different costs for their distance measure. Sensitivity analyses testing the effect of different cost specifications can be useful to affirm the validity of the choice of costs. For example, in a study by Mikolai et al,[177] sensitivity analyses were conducted with addition/subtraction operation weights adjusted to 0.5, 1.0 and 1.5. Substitution was also performed according to a constant substitution matrix, as well as a data driven approach whereby a cost was assigned according to the frequency of transitions between pairs of states in the observed data.[177]

Another limitation with optimal matching is that there are currently no clear guidelines on how to handle missing data.[62] In contrast to the LCA methods, whereby methods exist to incorporate the non-missing data from individuals with incomplete data into the

overall analysis, such methods can be problematic for optimal matching, as any observed dissimilarity may be due to comparing sequences of different lengths, rather than inherently dissimilar sequences.[178]

Once a (dis)similarity matrix is formed using an appropriate distance measure, the next step is to apply a type of clustering method to this matrix, to group individuals with similar distances into common trajectories.[179] The most commonly used approach is Ward's method for hierarchical clustering; this iteratively combines increasing numbers of sequences into a common trajectory whilst minimising the total distance within each trajectory at each iteration.[179]

As with the LCA trajectory derivation methods mentioned thus far, for sequence analysis too, the number of trajectories to model (in the cluster analysis step) requires prior specification by the analyst. Furthermore, in a study by Han et al,[179] it is advocated that a holistic approach to optimal model selection should be taken in sequence analysis, which considers theory, meaningfulness of classes (sequence index plots can be used to graphically assess the derived trajectory classes), and measures of the statistical quality of the classes (i.e., this is further in line with the approach to LCA methods discussed). However, as sequence analysis is not based on finite mixture models, the set of model fit indices for the LCA methods mentioned previously cannot be used here. Instead, other model fit indices specific to sequence analysis are available. For example, Hubert's C index[180] can be used, whereby higher values represent better fitting models.

Now, with LTA and sequence analysis described, the application and performance of these trajectory derivation methods, alongside other derivation methods from included studies in the systematic review of this thesis are reviewed in the next Section.

5.4 Critique of Trajectory Methodology from Systematic Review

In this Section, the systematic review from Chapter 3 is revisited, and the third objective is carried out: a critical appraisal of the trajectory derivation methodology used in the six included studies.

5.4.1 Different Trajectory Methods and Statistical Software Used

As summarised below in Table 13, two studies used sequence analysis to derive their trajectories of absence,[100,70] three used LCGA,[69,72,101] and a final study used LTA.[73]

Sequence analysis was used in the two relatively older studies (published in 2016[100] and 2018,[70] respectively), and resulted in derivation of the highest number of trajectories in the systematic review (eight[100] and nine,[70] respectively). Whilst the studies based on LCA methods of either LCGA or LTA were more recent (published from 2018-2023), and also resulted in fewer trajectories in their optimal models (either five or six).[69,72,73,101]

Table 13. Trajectory Derivation Methodology Summary from Six Included Studies of Systematic Review

First Author, Publication Year	Type of Absence Definition	Trajectory Derivation Method	Statistical Software Used	Number of Trajectories in Final Model	Model Building Process	Criteria Used to Select Optimal Model	Reported Performance Measures of Optimal Model
Pedersen, 2016	Categorical	Sequence Analysis	Stata (Package: SQ-Ados)	8	- Optimal matching used, then hierarchical cluster analysis applied	- None stated	- None reported
McLeod, 2018	Categorical	Sequence Analysis	Stata (Package: SQ-Ados)	9	- Optimal matching and cluster analysis originally used, but terminal RTW end state was deemed to be underweighted, hence bespoke decision rules used	- None stated	- None reported
Farrants, 2018	Continuous	LCGA	SAS (Procedure: Proc traj)	6	- Progressively increased number of classes until model fit no longer improved	- Lower BIC - Minimum trajectory prevalence ≥5% - Minimum average posterior probability ≥0.7	- Lowest trajectory prevalence was 6% - All classes had average posterior probability >0.9
Farrants, 2019	Continuous	LCGA	SAS (Procedure: Proc traj)	5	- Progressively increased number of classes until model fit no longer improved	- Lower BIC - Minimum trajectory prevalence ≥5% - Minimum average posterior probability ≥0.7	- Lowest trajectory prevalence was 8% - All classes had average posterior probability >0.9

Spronken, 2020	Categorical	LTA	Latent GOLD (Choice Module)	5	- Progressively increased number of classes until model fit no longer improved - All trajectory analyses run with 160 random starts and 250 iterations.	- Lower BIC - Minimum trajectory prevalence ≥5% - Clinically meaningful and distinct classes	- BIC = 148296 - Lowest trajectory prevalence was 9% - Entropy of 0.45 - Clinically relevant and interpretable classes
Rysstad, 2023	Continuous	LCGA	Stata (Package: traj)	6	- Progressively increased number of classes until model fit no longer improved - Also considered different polynomial functions from linear up to quintic, although strategy not clear	- Lower BIC - Minimum trajectory prevalence ≥5% - Minimum average posterior probability ≥0.7 - Distinct classes with narrow 95% confidence intervals - Clinically meaningful classes - Higher entropy values - Spaghetti plot of individual participant trajectories within a class also inspected	- BIC = -13493 - Lowest trajectory prevalence was 7.3% - All classes had average posterior probability > 0.95 - Entropy of 0.95 - Clinically relevant and interpretable classes

Abbreviations: BIC = Bayesian Information Criterion; LCGA = Latent Class Growth Analysis; LTA = Latent Transition Analysis; RTW= Return to Work.

Alongside these different methods, the statistical software used for trajectory derivation also varied, to include Stata for the two sequence analysis studies (using the package SQ-Ados), and Latent GOLD for LTA (with the Choice module). Then LCGA was performed with SAS (using Proc Traj) in the two Farrants et al studies,[69,72] and Stata (using the package traj) by Rysstad et al.[101]

5.4.2 Trajectory Reporting from Sequence Analysis Studies

In terms of how the trajectory derivation methodology was applied, beginning with the two sequence analysis studies, both Pedersen et al[100] and McLeod et al[70] considered optimal matching as a distance measure, followed by cluster analysis to assign individuals with similar sequences into common trajectories.

Whilst Pedersen et al[100] did ultimately use this approach (using Ward's linkage for hierarchical clustering - see Section 5.3.2 for a summary of this method), McLeod et al[70] deemed their derived trajectories to be illogical, as they considered this approach to underweight sequences that ended in a RTW state in their dataset. Therefore, McLeod et al[70] instead derived their own decision rules for trajectory grouping, based on: the terminal state of a trajectory (i.e., the final state in the ordered sequence of events over follow-up), the time to reach this terminal state, and the types of states that occurred prior to this terminal state. However, this bespoke decision rule was not described in any further detail, making it difficult to critique and compare to the Pedersen et al[100] study. Use of such a rule may well have been an important limitation, and the authors themselves stated: "other decision rules may have led to a different set of clusters, and the impact of these decisions are important to investigate in future research".[70(p154)]

Similarly, the Pedersen et al[100] study is also difficult to critique explicitly, as they did not provide any detail concerning which costs (if any) they used when performing optimal matching. Nor did either of these two studies consider any type of sensitivity analyses for other specifications of optimal matching, raising concerns about generalisability of findings.

In general, the reporting from these two sequence analysis studies was poor, and it was also concerning that no criteria were provided as to how the optimal trajectory models were chosen. Additionally, no summary of statistical measures was presented to demonstrate the fit of the final model, which made it difficult to critique the performance of these trajectory models.

5.4.3 Trajectory Reporting from LCA Studies

The standard of trajectory analysis reporting was improved in the four remaining studies based on LCA methods. All of these studies set out the criteria that they used to elect their optimal model. They then used these criteria in their model building strategies, whereby the number of trajectory classes were progressively increased until model fit no longer improved.

When applying LCGA, Farrants et al[69,72] used the same criteria in both of their studies: the optimal model was chosen based on having a lower BIC compared to other models. It was also required for all of the trajectories in this model to have at least 5% prevalence and an average posterior probability of at least 0.7.

Similarly, Spronken et al[73], in their application of LTA, also used the same optimal model selection criteria as Farrants et al,[69,72] except that there was no minimum requirement of average posterior probability. Instead, a different criterion in Spronken

et al's[73] study was that classes needed to be distinct and meaningful. Furthermore, Spronken et al[73] reported how many sets of random starts (160) and iterations (250) that they conducted their analysis with; this number of random starts is within the recommended thresholds proposed by Jung and Wickrama,[61] to avoid issues of local solutions when the log likelihood is being maximised (see earlier discussion regarding LCA model convergence issues in Section 5.2.4). Indeed, reporting the number of randoms starts and iterations is one of 16 items from the Guidelines for Reporting on Latent Trajectory Studies (GRoLTS) checklist (2016).[173] This checklist was published to encourage better quality reporting in studies that use mixture modelling (applicable to the LCA methods discussed in this Chapter, but not to sequence analysis) – see Appendix G for checklist in full.

The highest quality reporting in this review was performed by Rysstad et al,[101] in their application of LCGA. In fact, this study was stated to be reported in accordance with the GRoLTS checklist (2016).[173] Here, the same criteria for optimal model selection mentioned thus far across the Farrants et al[69,72] and Spronken et al[73] studies were used, and in addition, choice of the final model was informed by higher entropy values. Also, there was a more specific requirement for model estimates in each trajectory (average number of monthly sick leave days) to have narrow 95% CIs, and the trajectories of individuals within a class for the optimal model were visually explored through a spaghetti plot, to further ensure the meaningfulness of these classes.

Additionally, the Rysstad et al[101] study was the only study from this review to consider alternative polynomial functions in their model building approach (from linear up to quintic). Although, the exact strategy was not clearly stated (i.e., there was no mention of whether these polynomial functions were tested prior to, concurrently with, or after

testing of the optimal number of classes). Rysstad et al's[101] final six-class model contained one linear, three quadratic, and two cubic classes.

Then, after optimal LCA models had been chosen using the above criteria, in terms of the reported performance of such models, the two Farrants et al[69,72] studies provided few details. Whilst it was stated that the optimal model in both studies had an average posterior probability greater than 0.9 (substantially above the required 0.7), neither the BIC nor model entropy were reported, which are both key model performance items that should be reported according to the GRoLTS checklist (2016).[173] This restricts ability to compare the overall performance of Farrants et al's[69,72] optimal models to those of the other studies in this review.

Spronken et al[73] improved upon the reporting of Farrants et al,[69,72] and did provide a BIC and entropy value. Although there are no clear guidelines regarding entropy cut-offs (see earlier discussion in Section 5.2.6), Spronken et al's[73] reported entropy was 0.45, which is considerably lower than a suggested cut-off of ≥0.7 for good class separability (by Andruff et al).[169] This suggests that Spronken et al's[73] choice of optimal model could be improved upon.

Finally, Rysstad et al[101] also provided an entropy for their chosen model, which was deemed 'excellent' by the authors at 0.95, as well as a high average posterior probability that was >0.95 for all classes.

5.4.4 Review of Overall Standard of Trajectory Reporting

In summary, the trajectory derivation methods used in this review were varied, and concerns were identified regarding the quality of the trajectory reporting. High quality reporting is important, as this not only facilitates critical appraisal of a study and

comparison against other studies, but also promotes transparency and reproducibility of the analyses conducted.[62,173]

The exceptions were the two most recent studies of this review, Rysstad et al (2023)[101] and Spronken et al (2020).[73] Rysstad et al (2023)[101] in particular, exhibited the most comprehensive reporting, and largely adhered to the published GRoLTS checklist (2016)[173] for good practice in trajectory reporting.

In contrast, the two studies by Farrants et al from 2018[72] and 2019[69] were poorly reported. Whilst the general criteria used in the optimal model selection process were stated, the corresponding results of these criteria for their optimal models were not presented.

Then, the oldest studies of this review, McLeod et al (2018)[70] and Pedersen (2016),[100] exhibited the worst reporting. No details of any model selection criteria were provided.

Thus, the quality of trajectory reporting observed in this review may have been influenced by the timing of six included study publications. Given that the GRoLTS checklist[173] was published in 2016, and that trajectory methodology is still relatively novel, especially as applied to work absence research (as demonstrated by the few absence trajectory studies identified in this thesis' systematic review), it is plausible that the authors of the older studies may have been less informed about how to report trajectory methods well.

However, one commonality of poor reporting amongst all the studies in this review, was that the choice of trajectory derivation method used was generally not explained. In item 6a from the GRoLTS checklist[173] (see Appendix G), the checklist authors advocate that the chosen derivation method should be discussed, and recommend that in the case

where a LCGA derivation method is used (which they acknowledge is often chosen ahead of a GMM, due to computational issues with the latter), that ideally both a LCGA and GMM should be fitted and contrasted, as the different assumptions regarding heterogeneity within classes can lead to different results.

Pedersen[100] and McLeod et al[70] justified their use of sequence analysis for trajectory derivation ahead of non-individual centred longitudinal analysis approaches, but did not explain why this method was used instead of other individual-centred approaches. Of note, McLeod et al[70] did present lengthy discussion concerning the advances of sequence analysis in modelling repeated RTW measures, and focused on the benefits of this method over multi-state models, whilst they did not make a comparison against LCA methods. It may be plausible that both of these studies did not consider LCA methods due to the timing of their publications (2016[100] and 2018)[70] and given the limited use of LCA methods at that time in sickness absence research.

Whilst Rysstad et al (2023)[101] excelled in most areas of their trajectory reporting and given that they used LCGA reported according to the GRoLTS checklist, it was unexpected that they did not make any remark about GMM methods. As mentioned, in the GRoLTS checklist they explicitly recommend that GMM at least be considered if LCGA is used.

Similarly, the two studies by Farrants et al[69,72] provided limited justification of their choice of LCGA methodology.

In contrast, Spronken et al[73] did provide some justification of their choice of LTA, and compared this to LCGA, stating that "gradual RTW occurs in stepwise transitions rather than smooth increases",[73(p3)] and they deemed LTA to be better served for such a

purpose. Nonetheless, it was unusual, considering that one of the main benefits of LTA ahead of LCGA and GMM is that the former allows quantification of transitions across derived trajectories and the corresponding categorical states modelled, that Spronken et al[73] paid little attention to the transitions between the different RTW % states that they used in their study. They did make brief mention to the average number of transitions of RTW states for each of their five derived trajectories, however, the focus throughout their study was on the derived trajectories themselves. Indeed, their stated research objectives concerned trajectory derivation, and did not mention transitions between RTW states. Thus, it was not entirely clear if using LCGA or GMM, instead of LTA, might have sufficed for Spronken et al's[73] research objectives, perhaps through use of an appropriate polynomial functional form to allow for the stepwise transitions that Spronken et al[73] was seeking to model.

More generally, poor quality trajectory reporting has also been observed by other researchers. For example, in the process of creating the GRoLTS checklist,[173] the authors reviewed 38 studies where LCGA or GMM was applied to model posttraumatic stress. After reviewing these studies, the authors felt that "the way these models have been reported on in the past has not been as transparent and consistent as would be needed to produce trustworthy and replicable findings".[173(p464)]

Nguefack et al (2020),[62] also concurred with this finding in their narrative review of six different trajectory methods, that was based on epidemiological research as a whole (they reviewed two cross sectional methods, as well as four longitudinal methods that were all considered in this Chapter – LCGA, GMM, LTA and sequence analysis). The study authors stated that, "based on our review, the complete description of trajectory modelling techniques is often insufficient and lacks essential details."[62(p1217)] Nguefack

et al (2020)[62] posited that the poor reporting could be a consequence of manuscript space limitations, and suggested authors consider utilising appendices to ensure trajectory methods are properly reported.

5.4.5 Concluding Remarks

To conclude this narrative synthesis, three types of trajectory derivation method were observed in this review: sequence analysis, LCGA and LTA. However, GMM, an important trajectory derivation method that has been discussed in this Chapter, was not used in any of the studies included in this systematic review, nor was this method apparently considered by study authors.

Furthermore, the key finding of this critique of trajectory methodology, was that the standard of trajectory reporting as applied to derivation of sickness absence trajectories was generally poor. However, it was encouraging that the more recent studies reviewed did exhibit much improved reporting. Introduction of tools such as a checklist for higher quality trajectory reporting, and generally the increase in recent publications of absence trajectory studies may have led to this improvement.

With this general absence of clear reporting of model fit indices and statistics relating to the meaningfulness of derived trajectory classes, it was challenging to assess and compare the performance of the different trajectory derivation methods in this review.

Overall, better quality reporting of absence trajectory studies is needed. Additionally, studies should justify the use of their chosen trajectory derivation method, and in particular, studies that use LCGA should also consider GMM.

5.5 Reflection on LTA and Sequence Analysis

Finally, this Chapter closes with a discussion of the trajectory methods to take forward in this thesis.

After considering the suitability of LGCM, LCGA and GMM methods for this thesis' aims, as well as the strengths and limitations of these methods, it was decided to take all three methods forward for trajectory derivation analysis in Chapter 7 (as discussed in Sections 5.1 and 5.2).

Indeed, all three of these methods are related and types of GCM. LCGA and GMM are extensions of a LGCM (through allowing two or more trajectory classes to be identified in a study, rather than a single common trajectory), and a GMM is an extension of a LCGA (as it permits individuals to vary from the class-specific trajectory they are assigned to, unlike the latter method).

It remains to discuss whether or not to also use LTA and sequence analysis in addition to these three methods for absence trajectory derivation in this thesis.

LTA is semi-parametric and derived from mixture models, and consequently has some similarities to LCGA and GMM. However, LTA is not considered a competing model to LCGA and GMM, rather, an altogether different approach to describing change over time.[181] The key difference is that LTA adds another layer of complexity to the model, by allowing individuals to transition across the classes derived at each time point, thus these models are especially beneficial if theory dictates that individuals might not remain fixed in derived classes. In this thesis, the focus was on deriving absence trajectories and the additional complexity in simultaneously modelling for individuals to

transit across the derived trajectories was not deemed necessary. Therefore, LTA was not used in this thesis.

In contrast to LTA, sequence analysis has a different statistical foundation, and is completely non-parametric, thus is less similar to the LCA methods discussed in this Chapter. Indeed, this poses certain limitations, as compared to these LCA methods, there are less model fit indices available to evaluate the optimal sequence analysis models and it is currently unclear how best to incorporate missing data in analyses.[62]

Nonetheless, sequence analysis is a powerful statistical tool that can address the key aim of this thesis: identification of different subgroups of individuals at risk of long-term absence. Sequence analysis has been widely used for trajectory derivation, for example, its application in life course research has grown exponentially in the past decade and a half.[178]

However, comparing the results of sequence analysis to LCA methods can be challenging, especially due to the different foundations that these models are built upon, as well as the different choices that need to be made in the application of sequence analysis and the sensitivity of results to these choices (such as how the cost used in the distance measure process is specified). In this thesis' systematic review, both Rysstad et al[101] and McLeod et al[70] derived trajectories for individuals absent due to a MSK condition, using LCA and sequence analysis methodology, respectively. But in their discussion, Rysstad et al[101] stated that the McLeod et al[70] "study used a different modelling method (sequence analysis), making it difficult to directly compare it to our results".[101(p284)]

Though, a few studies have compared sequence analysis and LCA methods, and shown that the results across both types of trajectory derivation approach were similar.[86,177,179] In their comparison, Han et al (2017)[179] concluded that "our reasoning suggests that SA and LCA will most often lead to roughly the same typologies".[179(p337)] In this thesis' systematic review too, similarities in fast and slow RTW trajectories were observed across the six included studies, irrespective of the trajectory derivation method used.

Therefore, whilst sequence analysis is a useful trajectory derivation method, it was not used in this thesis as it was not expected that it would lead to vastly different results as compared to the optimal model obtained from the three LCA methods that were used.

5.6 Conclusion

In closing, in this Chapter the benefits of using individual-centred approaches to analysing longitudinal data were discussed, and a detailed summary of application of these methods was provided, based on three chosen trajectory derivation approaches for this thesis: LGCM, LCGA and GMM.

These methods will be revisited in Chapter 7 where they are used for derivation of work absence trajectories due to either a MSK or MH condition.

In the next Chapter, a study is first conducted on incidence rates of absence due to a MSK or MH condition. This Chapter examines trends in fit note incidence rates over recent years, and also serves as a preliminary step to performing trajectory derivation analyses in Chapter 7, as the cohort of individuals identified, with a first ever fit note due to a MSK or MH condition, are used to define the baseline cohort for the trajectory analysis.

Chapter 6. Study 1: Incidence Rates of Work Absence

6.1 Study Aim and Objective

The first original research study of this thesis involved a trends analysis, which aimed to establish incidence rates of work absence due to a MSK or MH condition, over the years from 2010 through to 2021.

A further objective was to examine whether these rates varied by sociodemographic characteristics.

6.2 Methods

6.2.1 Defining the Study Outcome

The primary outcome of this study was incidence of fit notes. In contrast to prevalence, which concerns existing and new occurrences of a particular event, incidence concerns new occurrences of the event only. The general structure of an incidence rate is defined as the number of new "events occurring in a defined population over a specified period of time (numerator), divided by the population at risk for that event over that time (denominator)".[182(p49)]

When calculating an incidence rate, there are two possible approaches one can take, depending on how the denominator is defined. The denominator can be defined either using a simple count of the number of persons at risk, or by using person-time units to factor time at risk during the specified period of follow-up into the calculation.

As an example, if there were 100 persons in the defined population at the start of follow-up, and 10 of these persons went on to have the specified event for the first time during a year's worth of follow-up, then the incidence rate for this year based on the

first method would simply be a count of the number of such events divided by the number of persons in the population: 10 events per 100 persons.

However, if it was also known that all 10 of these persons experienced these events after exactly three months of follow-up from baseline (i.e., each of these persons had been at risk for one quarter of the year of follow-up before the event occurred), and that the remaining 90 persons were all followed-up for a year without an event occurring, then the incidence rate using the second method could also be applied to take time-to-event into account. Here the numerator would remain the same (10 new events in the year of follow-up), but for the denominator, instead of counting all persons with a value of '1', the persons who experienced an event would instead be counted by multiplying the value of '1' by the proportion of time in the year that had passed before the event was experienced (i.e., 0.25 years). Thus the incidence rate would now be 10/((10x0.25)+90) = 10/92.5 = 108 events per 1000 person-years (correct to 3 decimal places).

In this study, incidence rates were calculated using methods based on both counting persons, and person-years, and results from each approach were compared. Rates based on numbers of persons at risk are presented per 10,000 persons, and rates based on person-time per 10,000 person-years.

The incidence rate outcome as applied to this study was defined as the rate at which new (incident) fit notes were issued (i.e., the first ever fit note for a particular person), over a fixed time period of one full calendar year, and whereby this first ever fit note was issued due to either a MSK or MH condition. A first ever fit note was determined through searching a person's lifetime history of fit notes received in the CPRD Aurum database.

From this definition the following three study outcomes were originally planned to be used: incidence rates of fit notes due to: a MSK condition only, a MH condition only, and due both a MSK and a MH condition simultaneously (defined as where there was at least one MSK and one MH consultation in the 14 days prior to the index fit note issue date).

However, the frequency of incident fit notes issued due to both MSK and MH conditions simultaneously was low (this may have been in part due to coding issues, whereby if a consultation occurred for multiple health conditions, only one health condition might have been chosen to be entered into the database). Therefore, there was insufficient data to look at co-occurrence of MSK and MH conditions, hence this outcome was not used.

Incidence rates were calculated using first ever fit note data from 1st January 2010 up to 31st December 2021. These dates allowed this study to incorporate the most recent data possible - the CPRD data request protocol was approved on 22nd December 2021, and the data manager, JB, extracted the February 2022 build of CPRD Aurum data, thus all fit note issuance data up to the end of 2021 was indeed available. Furthermore, choosing a start date of 1st January 2010 allowed a wide range of incidence rates to be calculated (twelve years' worth) for annual trend comparisons.

6.2.2 Defining the Study Population

Now the overall study population is defined, commencing with the numerator population (persons who experienced a new event), which is a subset of the denominator population (persons who were at risk of experiencing a new event).

The numerator population, for a particular year, consisted of persons that had been issued a first ever fit note during that year, and where this first ever fit note was due to a MSK or MH condition.

MSK conditions were defined using Read/SNOMED codes (n=723) from the MSKCOM Keele based study (CPRD study reference 20_000105)[125] that also used CPRD Aurum. MSK conditions in this study were defined as the following six conditions:

- Osteoarthritis
- Inflammatory MSK

Or the most common regional pain:

- Back pain
- Knee pain
- Hip pain
- Hand/wrist pain

These code lists are publicly available on the Keele Research Repository (https://doi.org/10.21252/878s-x990). After receiving clinical expertise from members of our CPRD data request protocol study team, it was determined that these definitions would cover the most common definitions of MSK conditions in issued fit notes.

MH conditions were also defined with the Read/SNOMED codes (n=189) from the MSKCOM study[125] (code list also available publicly from https://doi.org/10.21252/878s-x990), as:

- Depression

- Anxiety
- Stress

These three conditions were also deemed to cover the most commonly certified MH conditions.

However, reason for fit note issuance data is not available in CPRD. Therefore, an assumption was made using medical consultation history data (which is available in CPRD), whereby if there was a consultation for a MSK or MH condition that occurred in the two weeks prior to or on the date of fit note issuance, then the fit note could be attributed to this same MSK or MH condition.

In the original CPRD data request protocol (Appendix E), it was first planned to consider MSK or MH conditions that occurred ± two weeks of the index fit note. This was later restricted to MSK or MH conditions that occurred in the two weeks prior to fit note issuance only (Appendix F), as further discussion with GPs suggested that consultation codes recorded after fit note issuance were less likely to be the reason for the fit note.

Four other inclusion and exclusion criteria to define the numerator population were also applied, as detailed below in Table 14.

Table 14. Inclusion and Exclusion Criteria Used to Define Numerator Population for Incidence Rates Analysis

Inclusion	Exclusion
- A first ever fit note during the year of incidence rates calculation (covering 2010-2021), whereby there was also a MSK or MH consultation that occurred in the two weeks prior to this first ever fit note - Persons aged 16 to 66 years - At least 2 years prior registration with GP Practice	- Persons that died, de-registered, or had last collection date before the end of the year of the incidence rates calculation - Persons who were not registered with an English practice

Abbreviations: GP = General Practitioner; MH = Mental Health; MSK = Musculoskeletal

First, in line with the earlier systematic review of this thesis, in order to try to identify only economically active persons, an age restriction of between 16 and 66 years (the current UK SPA) was used. In CPRD, although date of birth data is not provided (to reduce the likelihood of researchers being able to identify persons in the dataset), year of birth data is available. Hence, in this study, age was calculated separately for each year from 2010 to 2021, by subtracting the person's year of birth from the chosen year of the incidence rate calculation.

Second, to ensure that each person had been with their GP practice for sufficient time, it was also required that persons were registered with their practice for at least two years by the end of the year of incidence rates calculation. For example, for 2010 incidence rates, in order to be eligible for study inclusion, persons were required to have had a registration date prior to or on 31st December 2008.

Third, as there was only a small number of people who did not have a complete year of follow-up data for each year, this group was excluded. Therefore, persons that had died,

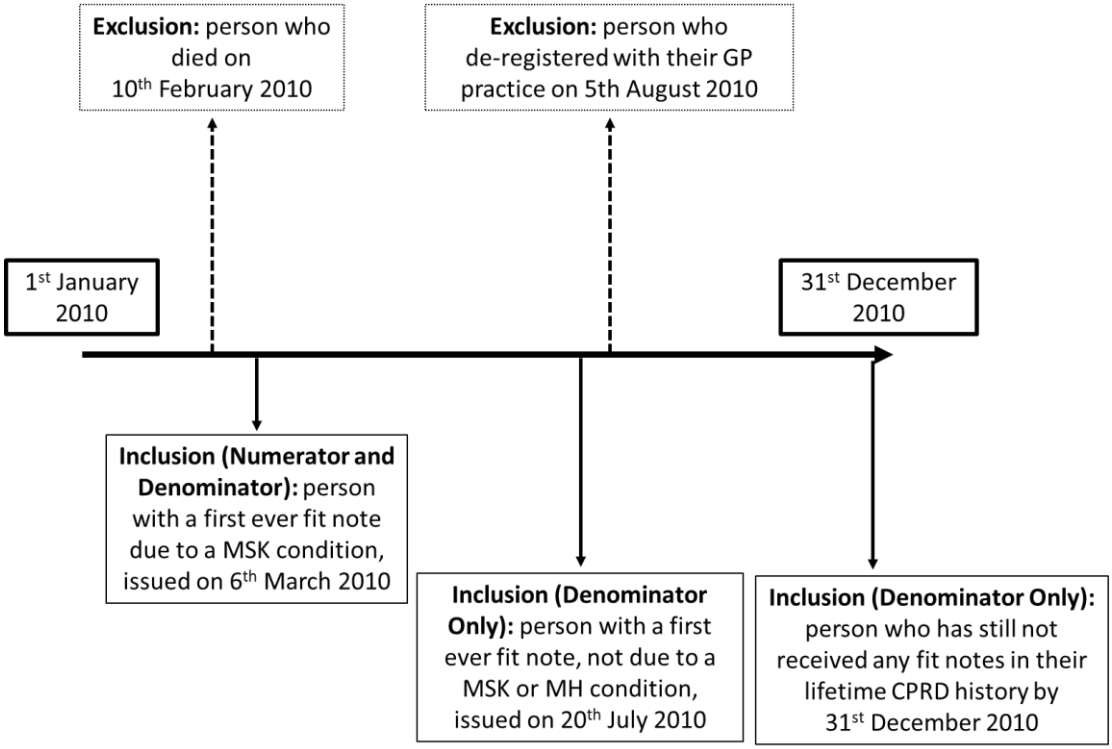
de-registered from their practice, or had last collection date during or before the year of incidence rates calculation were excluded.

Fourth, this study was based on an English population. The GP practice data from the CPRD Aurum database build used in this study (Section 4.4), predominantly was from practices in England (99.0% coverage). Therefore, the few persons that were registered with non-English GP practices were also excluded.

To define the denominator population, the same inclusion/exclusion criteria as above was applied, except that the requirement of a first ever fit note due to a MSK or MH condition during the year of incidence rates calculation was removed. Thus, persons who were yet to be issued a first ever fit note in the year of incidence rates calculation were included.

Furthermore, in the denominator population, persons who did receive a first ever fit note during (but not before) the year of incidence rates calculation, and where this fit note was not due to a MSK and/or MH condition, were also still included. This was because these people would potentially have been at risk of having been issued a first ever fit note due to a MSK and/or MH condition from the start of the specified year of incidence rates calculation (up to the point of the issued fit note). An example to illustrate this, alongside other reasons for inclusion/exclusion, is shown in Figure 10 below.

Figure 10. Diagram Showing Several Inclusion/Exclusion Examples for 2010 Incidence Rates Analysis



Abbreviations: GP = General Practitioner; MH = Mental Health; MSK = Musculoskeletal

6.2.3 Identification of Fit Notes

In order to ensure that there would be a sufficient number of individuals available for this study (as well as studies 2 and 3), a feasibility count was performed, to approximate the number of individuals with a new fit note in a given time period. Before the feasibility count process is described, the approach used to define fit notes is detailed.

Firstly, a list of Medical Code IDs relating to fit notes was devised using Read/SNOMED codes (Appendix H, Table H.1). CPRD converts the GP entered Read/SNOMED codes into their own medical codes (using their own coding system), and a tool known as Code Browser is available to enable users to convert between the standard and CPRD codes.

To devise this final set of fit note codes, an initial set of Read codes that had been previously derived by Keele researchers to identify fit notes (Read codes starting with: 9D1*, 9D2*, 13JJ*, and 13JX), was searched for in Code Browser (note: an asterisk ('*') represents truncation, hence, for example, 9D1* searches for terms such as: 9D1, 9D1A, 9D1B etc).

Running this search yielded 30 distinct Medical Code IDs. Then, the SNOMED Concept IDs corresponding to these 30 Medical Code IDs were searched for directly in Code Browser too (to identify any possible extra Medical Code IDs related to the same SNOMED Concept ID). This was a necessary extra step due to the structural system changes that followed implementation of SNOMED from Read Codes in 2018. This process identified 9 new Medical Code IDs, thus resulting in a final set of 39 Medical Code IDs that related to fit notes.

6.2.4 Feasibility Count

Next, using this list of 39 Medical Code IDs (Appendix Table H.1), a feasibility count was performed by the data manager, JB, in order to gain an understanding of the number of individuals with new fit notes, both in former times (the years 2010-2014 were chosen), and more recently (2019 was chosen).

Additional criteria applied by JB required that the individuals had at least two years prior registration (in line with study inclusion criteria from Table 14), and an age ≥ 16 years old (to correspond to adult participants that were at least of working age). These counts of first fit notes are shown in Table 15.

Table 15. Feasibility Count of First Fit Notes Due to a MSK or MH Condition

Year	First Fit Notes, n	Estimate of First Fit Notes Due to a MSK Condition, n	Estimate of First Fit Notes Due to a MH Condition, n
2010-2014	1,675,025	294,579	696,661
2019	723,484	127,236	300,905

Abbreviations: MH = Mental Health; MSK = Musculoskeletal

As mentioned in Section 6.2.2, reason for fit note issuance data is not available in the CPRD. Thus, an approach was used to estimate feasibility counts of first fit notes due to a MSK or MH condition. Using the most recent fit note issuance data from NHS Digital (from March 2021, the most recent file at the time the data request protocol was submitted), it was observed that 17.6% of total fit notes issued in England with an ICD Chapter 10 were due to 'Diseases of the musculoskeletal system and connective tissue', and 41.6% due to 'Mental and behavioural disorders'.[183] These ratios were applied to the feasibility count of first fit notes, to provide estimates of a count differentiated by MSK and MH conditions (Table 15).

In summary, feasibility counts suggested approximately 300,000 individuals from 2010-2014 (or 60,000 per year) with a first fit note due to a MSK condition, which increased to approximately 130,000 per year in 2019.

Similarly, feasibility counts suggested approximately 700,000 individuals from 2010-2014 (or 140,000 per year) with a first fit note due to a MH condition, which increased to approximately 300,000 per year in 2019.

Note: since the original feasibility count process was run, the set of 39 Medical Code IDs for identifying fit notes in CPRD (Appendix Table H.1) was further revised, by removing 12 Medical Code IDs that related to duplicate or non-issued fit notes. In the

final list there were 27 Medical Code IDs that related to fit notes in CPRD, as presented in Appendix Table H.2.

6.2.5 Categorising Fit Note Codes into 'Not Fit' and 'Maybe Fit For Work'

Prior to performing analyses, data cleaning was required to de-duplicate the numerator data in instances where a person received more than one type of fit note on the date of their first ever fit note.

Thus, the final set of 27 Medical Code IDs used to identify fit notes in the CPRD Aurum dataset (Appendix Table H.2) needed to be categorised into 'not fit for work', or one of the 'may be fit for work' options ('a phased return to work', 'amended duties', 'altered hours', or 'workplace adaptations').

AL and GWJ jointly used their best judgement to decide how to perform this categorisation of Medical Code IDs. A summary of how many fit notes ended up in each such category, as based on a near final numerator dataset of first ever fit notes due to a MSK and/or MH condition over 2010-2021, is shown below in Table 16.

Table 16. Summary of Fit Note Categorisation of the 27 Medical Code IDs Used to Define Fit Notes[a]

Medical Code ID	Term	Chosen Fit Note Categorisation	Broader Categorisation of 'Not Fit' vs 'Maybe Fit'	Frequency of Fit Notes, n	% of Total Fit Notes
1653351000000110	eMED3 (2010) new statement issued, not fit for work	Not fit for work	Not fit for work	349,941	68.73
11561000000115	Med3 certificate issued to patient	Not fit for work	Not fit for work	60,231	11.83
11551000000118	MED3 - doctor's statement	Not fit for work	Not fit for work	48,522	9.53
1653661000000110	eMED3 (2010) new statement issued, may be fit for work	Maybe fit for work	Maybe fit for work	24,419	4.80
1653921000000110	MED3 (2010) issued by hand, not fit for work	Not fit for work	Not fit for work	16,264	3.19
250873012	Unfit for work	Not fit for work	Not fit for work	1,886	0.37
1653961000000110	MED3 (2010) issued by hand, may be fit for work	Maybe fit for work	Maybe fit for work	1,365	0.27
34201000000114	Med3 certificate issued - back to work	Exclude	Exclude	1,350	0.27
11621000000112	MED5 - doctor's special statement	Not fit for work	Not fit for work	1,199	0.24
11591000000114	Med3 certification status	Not fit for work	Not fit for work	1,004	0.20
11631000000114	MED5 issued to patient	Not fit for work	Not fit for work	894	0.18

1769621000006110	MED3 (2010) certificate issued to patient	Not fit for work	Not fit for work	887	0.17
250932010	Time off work	Not fit for work	Not fit for work	551	0.11
1769661000006110	MED3 (2010) certificate issued - recommend amended duties	Maybe fit for work - Amended work duties	Maybe fit for work	375	0.07
1769651000006110	MED3 (2010) certificate issued - recommend altered hours	Maybe fit for work - Altered hours	Maybe fit for work	72	0.01
1769671000006110	MED3 (2010) certificate issued - recommend workplace adaptation	Maybe fit for work - Workplace adaptations	Maybe fit for work	62	0.01
1769641000006110	MED3 (2010) certificate issued - recommend phased return to work	Maybe fit for work - Phased RTW	Maybe fit for work	58	0.01
1156491000000110	MED5 certificate requested	Not fit for work	Not fit for work	23	0.00
11661000000116	MED5 status	Not fit for work	Not fit for work	6	0.00
1148561000000110	MED5 statement requested	Not fit for work	Not fit for work	5	0.00
735851000000113	Benefits agency reports unfit for work	Exclude	Exclude	5	0.00
11641000000117	MED5 - issued to patient	Not fit for work	Not fit for work	0	0.00
7968061000006110	Med3 certificate issued to patient	Not fit for work	Not fit for work	0	0.00
12487881000006100	MED3 issued to patient	Not fit for work	Not fit for work	0	0.00

12487871000006100	MED3 issued - back to work	Exclude	Exclude	0	0.00
7999471000006110	Med3 certificate issued - back to work	Exclude	Exclude	0	0.00
4536821000006110	Amount of time off work	Not fit for work	Not fit for work	0	0.00

[a] Based on a near final numerator dataset of n=509,119 data rows (containing duplicate entries per person) of first ever fit notes issued over 2010-2021, due to a MSK and/or MH condition, pertaining to n=399,565 persons

The majority of the data presented (98.09%) was represented by just five Medical Code IDs, with four of these relating to terms that were categorised as ‘not fit for work’:

- ‘eMED3 (2010) new statement issued, not fit for work’ (68.73% of fit notes)
- ‘Med3 certificate issued to patient’ (11.83%)
- ‘MED3 - doctor's statement’ (9.53%)
- and ‘MED3 (2010) issued by hand, not fit for work’ (3.19%)

The fifth of these most commonly occurring Medical Code IDs was categorised as ‘maybe fit for work’: ‘eMED3 (2010) new statement issued, may be fit for work’ (4.80%).

The remaining twenty-two Medical Code IDs occurred with low frequency (with each one representing less than 0.4% of the total data).

The specific types of ‘maybe fit for work’ fit note options of ‘amended work duties’, ‘altered hours’, ‘workplace adaptations’, or ‘phased return-to-work’ each only occurred with at most 0.07% of the total fit notes. Thus, due to these small numbers, it was decided not to use this level of categorisation, but to focus more broadly on ‘not fit’ or ‘maybe fit for work’ in this study.

Also, three fit note terms were excluded at this stage as they contained the words ‘back to work’ (two of these Medical Code IDs had the same term of: ‘Med3 certificate issued - back to work’, and the third: ‘MED3 issued - back to work’). These fit notes were back dated, and thus it was not possible to identify the true start dates of these fit note using the available CPRD data, rather only when the back dated fit notes were issued. Two of these three terms were not used in the dataset being analysed, and the other only

represented 0.27% of the total data, therefore these exclusions did not result in a significant loss of data.

The Medical Code ID relating to the term ‘Benefits agency reports unfit for work’ was also excluded, as this corresponded to unemployed persons, which was against the inclusion/exclusion criteria of this study. There were only five such instances of this term in the data, which was also not a significant loss of data.

Thus, after these exclusions, there was a final set of 23 Medical Code IDs that were used to search for fit notes in this study, and a further 4 of these codes did not retrieve any results.

Finally, in instances where a person received multiple first ever fit notes on the same date, that were categorised as both ‘not fit for work’ and ‘maybe fit for work’, only the ‘not fit for work’ fit note was retained in the de-duplication process. This was done on the premise that ‘not fit for work’ fit notes are the most commonly occurring type of fit note, as demonstrated in the Systematic Review by Dorrington et al (2018),[50] and the latest NHS Digital fit note issuance summary data (from April 2021 – September 2023).[38]

After the data was fully de-duplicated, to result in a final numerator dataset containing one fit note row per person, analyses were performed, as described in the next Section.

6.2.6 Analysis Plan

Incidence rates of fit note issuance due to a MSK or MH condition were analysed separately, for each year from 2010 to 2021, using methods based on both counting persons as well as person-years.

Trends in these yearly incidence rates were evaluated descriptively, separately for fit notes due to MSK and MH conditions. Trends between fit note incidence due to a MSK or MH condition were also compared descriptively, and through incidence rate ratios (IRRs).

Furthermore, all incidence rates were also stratified by the following sociodemographic characteristics: age, sex, and geographical region.

The continuous variable age was categorised into the following groupings: 16-25 years, 26-35 years, 36-45 years, 46-55 years, and 56-66 years.

Sex originally consisted of the groupings: male, female, and indeterminate (where it was not known whether the person was male or female). Persons with 'indeterminate' sex were ultimately excluded from all reporting due to low counts (always < 0.04% of the numerator population for a given incidence rate year, for incident fit notes due to either a MSK or MH condition).

The region variable used was based on ONS Region for English practices. After already excluding GP practices based in Northern Ireland (as mentioned in Section 6.2.2), there were nine regional groupings remaining for GP practices based in England: Northeast, Northwest, Yorkshire and the Humber, East Midlands, West Midlands, East of England, London, Southeast, and Southwest. These groupings were all retained.

Finally, although it was planned for the percentage of total incident fit notes issued by type of fit note ('not fit for work', or one of the 'may be fit for work' options: 'a phased return to work', 'amended duties', 'altered hours', or 'workplace adaptations') to be reported, there were insufficient counts in the different types of maybe fit for work options (as mentioned in Section 6.2.5). Hence a broader comparison was performed,

comparing only issuance of ‘not fit’ against ‘maybe fit for work’ incident fit notes for each year.

All analyses were performed using Stata MP version 17.0.

6.2.7 Missing Data

Generally, in the CPRD Aurum dataset used for this study, most of the intermediary variables used to define the study population, the stratification variables, and the outcome of this study had full data present (for example, age, sex, registration start date).

GP practice geographical region was one of the few variables with some level of missing data, but as this was not too large it was decided to exclude persons with missing data from all analyses (during the data cleaning process, n=3,757 persons were lost as a result of this, out of n=397,344 numerator persons that remained at that stage of data cleaning).

Furthermore, although there were some individuals who did not have a complete year of follow-up data for the incidence rates calculation, these were few in number and excluded (as mentioned earlier in Section 6.2.2).

However, one significant area of missing data concerned fit note duration. It was originally planned for median length of incident fit note duration to be reported by year in this study. But, of the n =378,943 persons that were in the final de-duplicated numerator dataset of incident fit notes due to a MSK and/or MH condition from 2010-2021, only n=1,645 (0.43%) of these persons had a fit note duration reported in the CPRD Aurum dataset, which was insufficient to perform such analyses. This missing data also presented challenges to the original plans for studies two and three of this

thesis, which is addressed in further detail in Section 7.2.3. The reason for this missing fit note duration data is explained later in this thesis' Discussion Chapter, in Section 9.2.2.

6.3 Results

6.3.1 Overall Incidence Rates

Overall incidence rates for fit notes due to a MSK or MH condition, are shown in Table 17 and Table 18, respectively. A visual representation of the trends in these incidence rates is also provided in Figure 11.

Table 17. Summary of Overall Incidence for Fit Notes Due to a MSK Condition

Year	Numerator, n	Denominator, n	Incidence Rate, per 10,000 persons (95% CI)	Incidence Rate, per 10,000 person-years (95% CI)
2010	11,911	6,798,764	17.52 (17.21, 17.84)	17.55 (17.24, 17.87)
2011	11,293	6,835,203	16.52 (16.22, 16.83)	16.55 (16.25, 16.86)
2012	12,856	6,906,955	18.61 (18.29, 18.94)	18.65 (18.33, 18.97)
2013	16,346	6,757,873	24.19 (23.82, 24.56)	24.25 (23.88, 24.62)
2014	17,289	6,778,359	25.51 (25.13, 25.89)	25.57 (25.20, 25.96)
2015	16,681	6,858,785	24.32 (23.95, 24.69)	24.38 (24.02, 24.76)
2016	15,100	7,034,439	21.47 (21.13, 21.81)	21.52 (21.18, 21.86)
2017	13,942	7,224,947	19.30 (18.98, 19.62)	19.34 (19.02, 19.67)
2018	13,222	7,359,466	17.97 (17.66, 18.28)	18.01 (17.70, 18.32)
2019	12,512	7,416,564	16.87 (16.58, 17.17)	16.91 (16.62, 17.21)
2020	8,031	7,341,391	10.94 (10.70, 11.18)	10.96 (10.72, 11.20)
2021	8,790	7,358,263	11.95 (11.70, 12.20)	11.97 (11.72, 12.22)

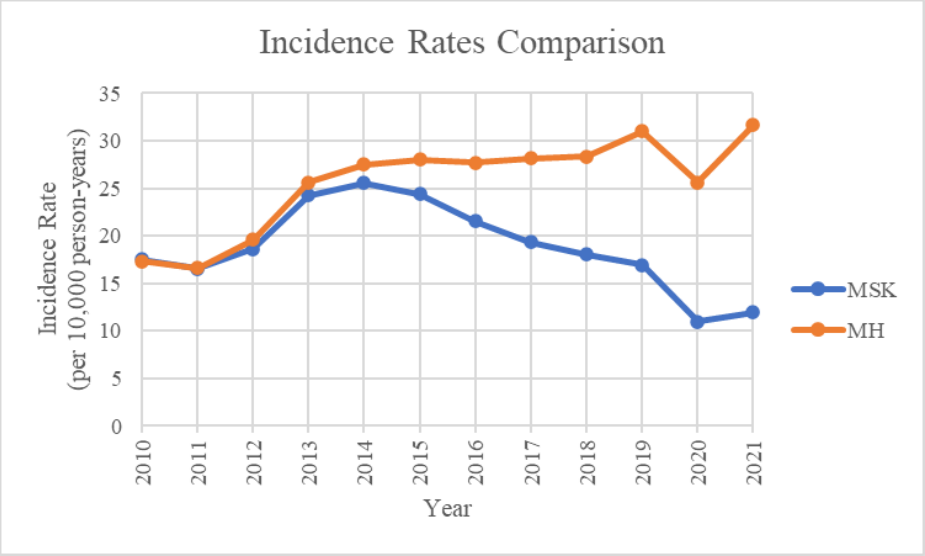
Abbreviations: CI = Confidence Interval; MSK = Musculoskeletal

Table 18. Summary of Overall Incidence for Fit Notes Due to a MH Condition

Year	Numerator, n	Denominator, n	Incidence Rate, per 10,000 persons (95% CI)	Incidence Rate, per 10,000 person-years (95% CI)
2010	11,737	6,798,764	17.26 (16.95, 17.58)	17.29 (16.98, 17.61)
2011	11,340	6,835,203	16.59 (16.29, 16.90)	16.62 (16.31, 16.93)
2012	13,491	6,906,955	19.53 (19.21, 19.87)	19.57 (19.24, 19.90)
2013	17,290	6,757,873	25.59 (25.21, 25.97)	25.65 (25.27, 26.03)
2014	18,582	6,778,359	27.41 (27.02, 27.81)	27.49 (27.09, 27.88)
2015	19,181	6,858,785	27.97 (27.57, 28.36)	28.04 (27.65, 28.44)
2016	19,442	7,034,439	27.64 (27.25, 28.03)	27.71 (27.32, 28.10)
2017	20,302	7,224,947	28.10 (27.72, 28.49)	28.17 (27.78, 28.56)
2018	20,823	7,359,466	28.29 (27.91, 28.68)	28.36 (27.98, 28.75)
2019	22,949	7,416,564	30.94 (30.54, 31.35)	31.02 (30.62, 31.42)
2020	18,776	7,341,391	25.58 (25.21, 25.94)	25.62 (25.26, 25.99)
2021	23,227	7,358,263	31.57 (31.16, 31.97)	31.63 (31.23, 32.04)

Abbreviations: CI = Confidence Interval; MH = Mental Health

Figure 11. Comparison of Incidence of Fit Notes Due to a MSK and MH Condition



Abbreviations: MH = Mental Health; MSK = Musculoskeletal

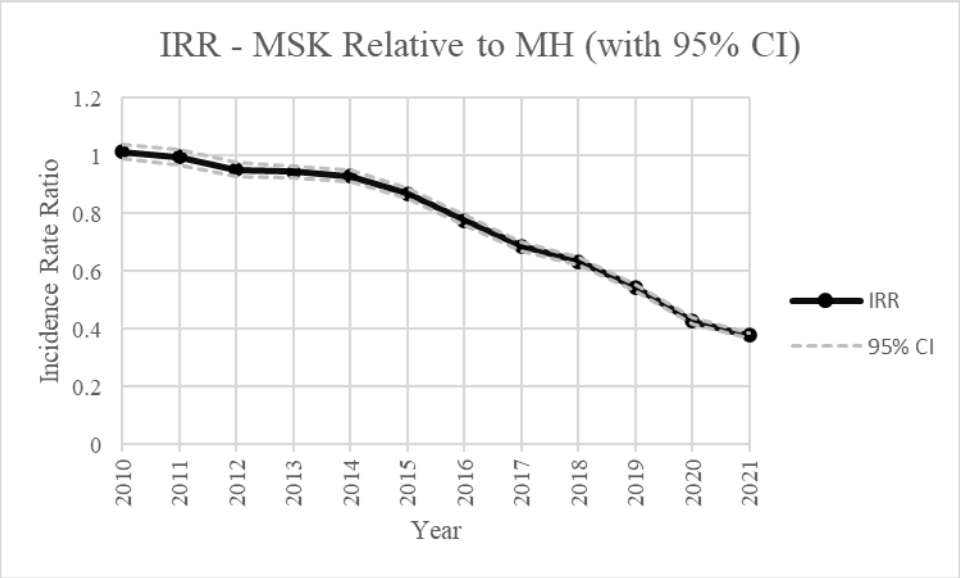
Firstly, for all fit note incidence rates (across all years and due to either a MSK or MH condition), there was negligible difference between these rates across the count of persons and count of person-years methods.

Incidence rates of fit notes for a MSK and MH condition started off similar in 2010 (17.55 and 17.29 per 10,000 person-years, respectively), and then both decreased slightly in 2011, before increasing steadily until 2014.

However, from 2014 onwards there was a difference observed, with fit note incidence due to a MH condition steadily increasing until 2021, whilst incidence due to a MSK condition was consistently decreasing. The only exception to these trends was an abrupt decrease for both MSK and MH fit note incidence in 2020.

The IRR of fit notes due to a MSK compared to MH condition quantified these descriptive trends (Figure 12). The IRR was not statistically significant and close to 1 in 2010 and 2011. From 2012 the IRR became and remained statistically significant up to 2021, and showed a continual year-on-year decrease, ending at 0.38 (95% CI: 0.37,0.39) in 2021.

Figure 12. Incidence Rate Ratios of Fit Notes Due to a MSK Relative to a MH Condition



Abbreviations: CI = Confidence Interval; IRR = Incidence Rate Ratio; MH = Mental Health; MSK = Musculoskeletal.

6.3.2 Incidence Rates by Age

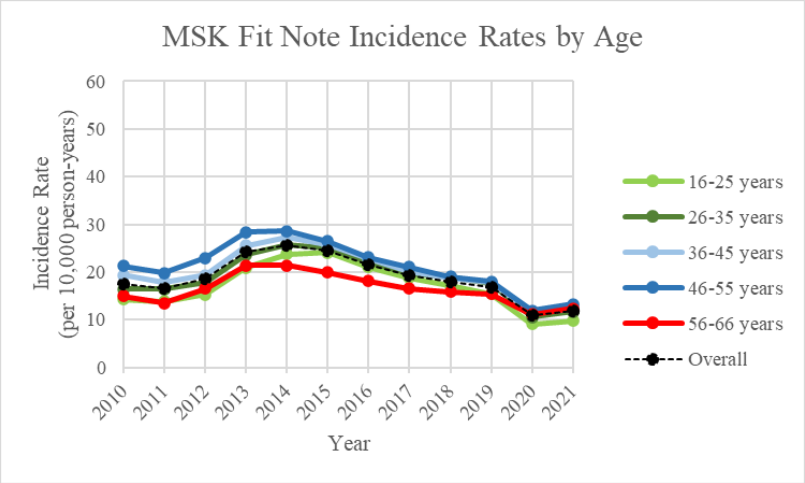
Incidence, stratified by age is visually depicted for fit notes due to a MSK condition in Figure 13, and due to a MH condition in Figure 14.

For fit note incidence due to a MSK condition, age trends were present, albeit less pronounced than for incidence due to a MH condition. Older age groups generally exhibited higher MSK incidences than younger age groups, across all years. The only exception to this was the oldest age group of 56-66 years, which had a consistently low incidence. However, these differences in incidences by age groups for a MSK condition became progressively more negligible in more recent years, to the point of there being almost no age difference from 2019 to 2021.

In contrast, for MH incident fit notes, the younger age groups generally always had higher and increasing incidence rates compared to older age groups. From around 2013 onwards, the incidences of the three oldest age groups stayed relatively stable until

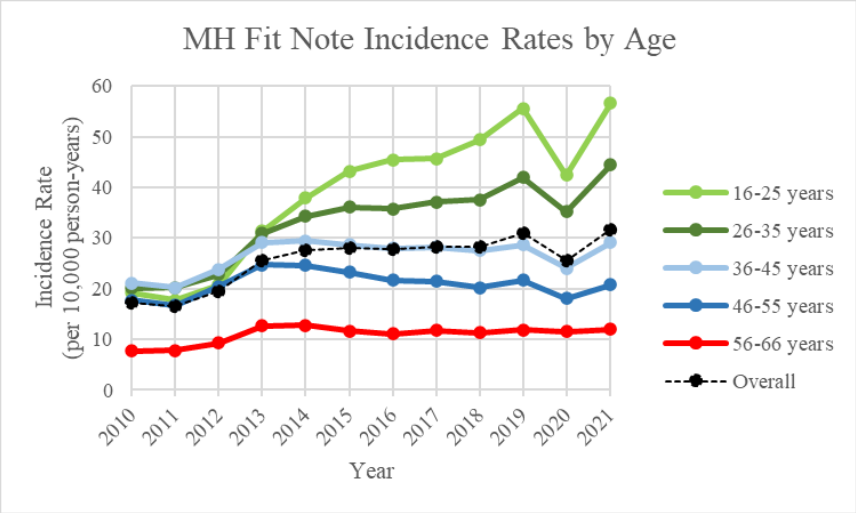
2021, whereas incidence for the 26-35 years group was steadily increasing over this time period, and incidence for the 16-25 years group was increasing the most quickly.

Figure 13. Incidence of Fit Notes Due to a MSK Condition by Age



Abbreviations: MSK = Musculoskeletal

Figure 14. Incidence of Fit Notes Due to a MH Condition by Age

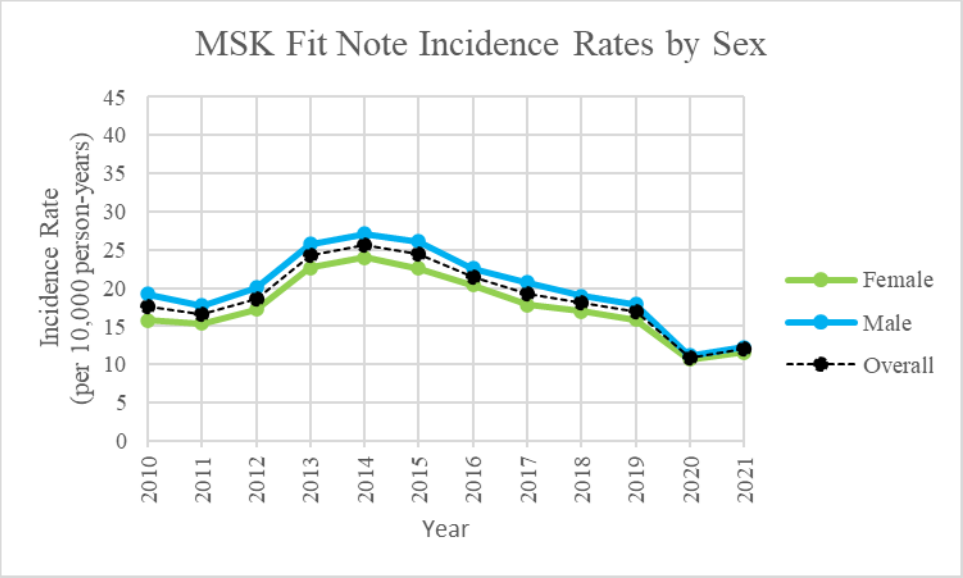


Abbreviations: MH = Mental Health

6.3.3 Incidence Rates by Sex

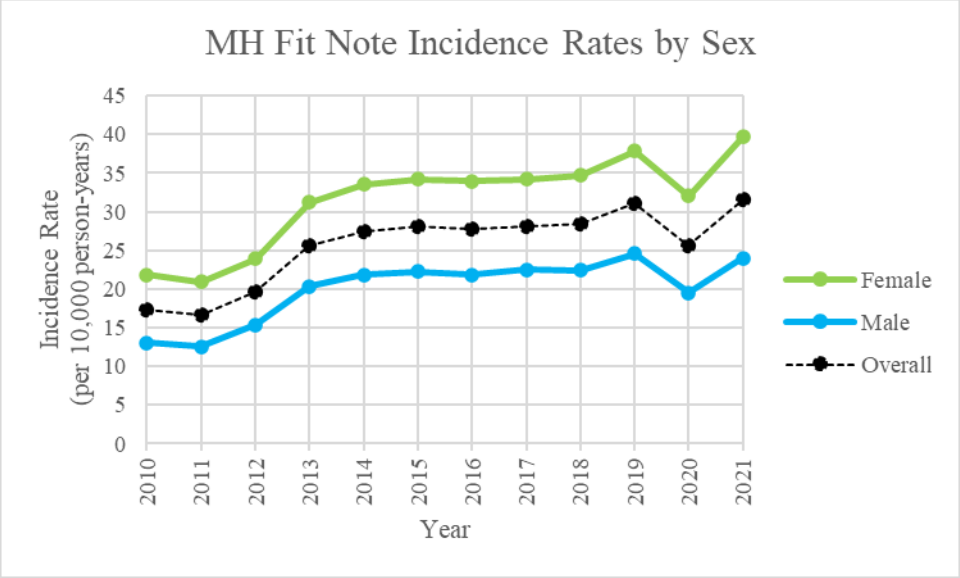
Stratifying incidences by sex, there were again differences in incidences of fit notes for MSK and MH conditions, albeit MSK condition sex differences were again less pronounced (Figure 15 and Figure 16, respectively).

Figure 15. Incidence of Fit Notes Due to a MSK Condition by Sex



Abbreviations: MSK = Musculoskeletal

Figure 16. Incidence of Fit Notes Due to a MH Condition by Sex



Abbreviations: MH = Mental Health

For MSK conditions, fit note incidence was higher for males than females all throughout. This sex inequality had almost completely levelled off by later years, especially in 2020 and 2021 where the IRR for females compared to males reached 0.95.

Whereas the reverse trend was observed for fit notes due to a MH condition, with incidence higher for females than males. There was a sustained difference in rates between females and males too, as demonstrated by an IRR that was always >1.52 for all years (comparing incidence of females to males).

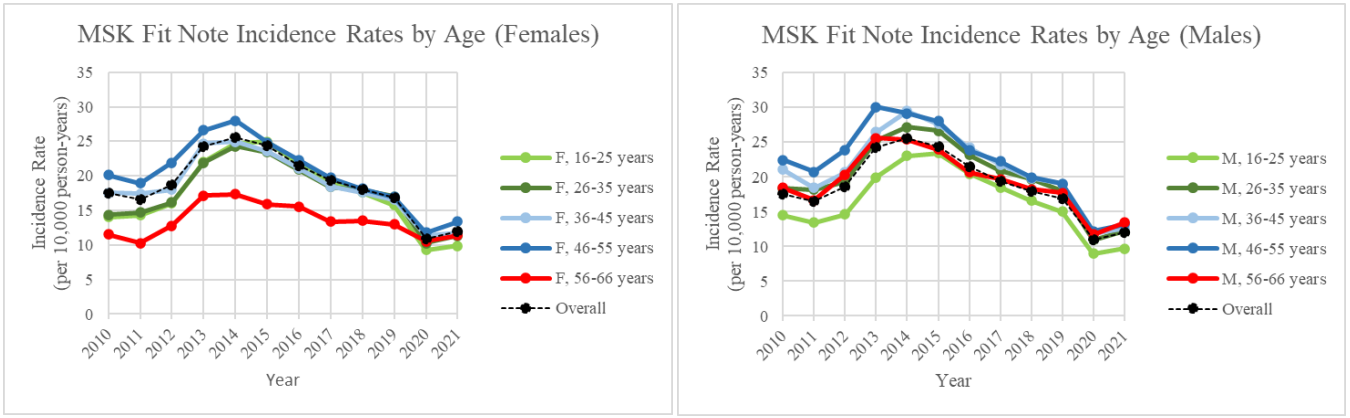
6.3.4 Incidence Rates by Age and Sex

Next, a stratification was performed on incidences by both age and sex. Age-sex incidences for fit notes due to a MSK condition are shown in Figure 17, and due to a MH condition in Figure 18.

Incidences by age for fit notes due to a MSK condition were generally similar for females and males, with the only differences being that the oldest males (56-66 years) and those aged 26-35 years had slightly higher incidences than their female counterparts.

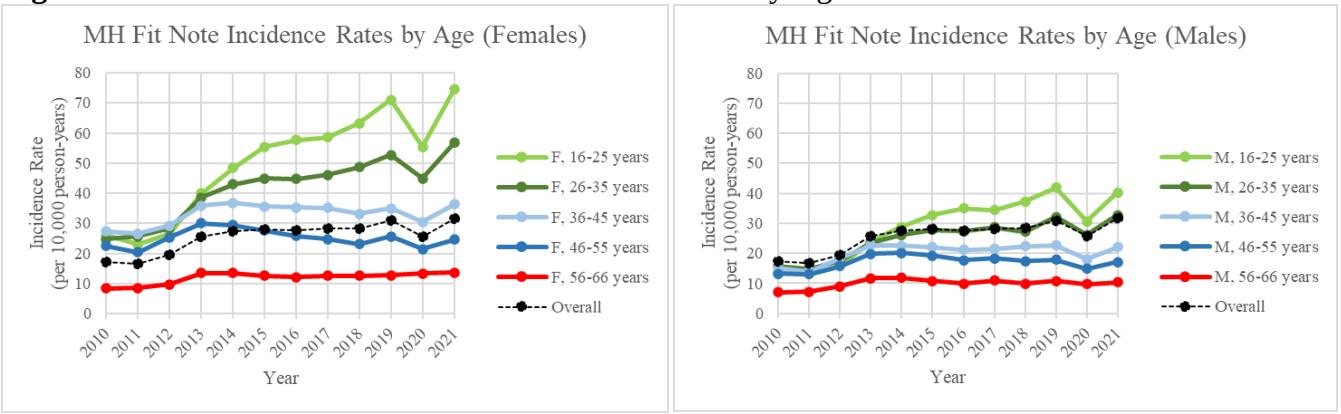
However, for fit notes due to a MH condition, more marked age differences were observed by sex. Females had higher incidences than males for the same age group, except for the oldest age group where incidence was approximately equal. Younger females in particular, had substantially higher incidences than younger males. The youngest female group (16-25 years) had the highest incidences over time that were also increasing most quickly. At the end of follow-up in 2021, 74.47 incident fit notes per 10,000 person-years were issued for this young female subgroup, compared to 40.22 per 10,000 person-years for males aged 16-25 years.

Figure 17. Incidence of Fit Notes Due to a MSK Condition by Age and Sex



Abbreviations: MSK = Musculoskeletal

Figure 18. Incidence of Fit Notes Due to a MH Condition by Age and Sex

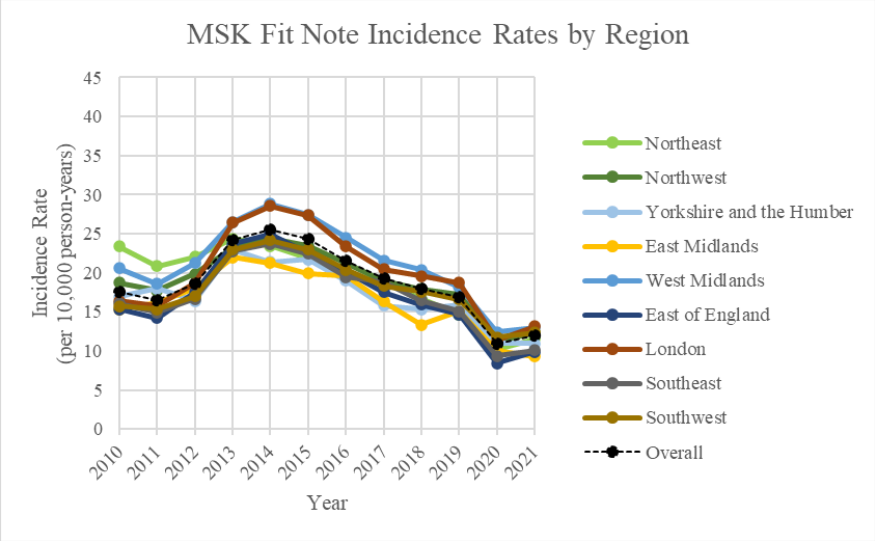


Abbreviations: MH = Mental Health

6.3.5 Incidence Rates by Geographical Region

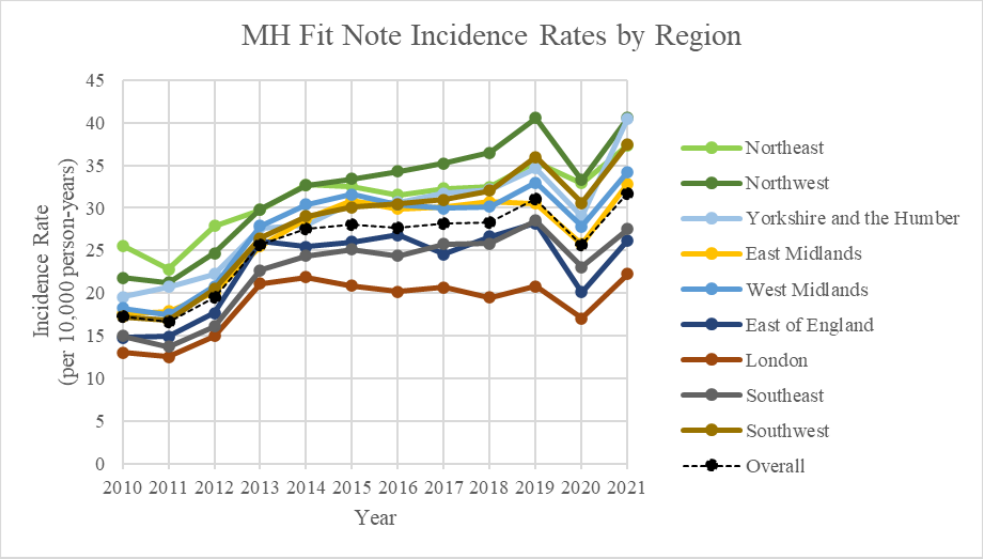
The final stratification of fit note incidence results was performed by region of GP practice (Figure 19 and Figure 20, for MSK and MH conditions, respectively).

Figure 19. Incidence of Fit Notes Due to a MSK Condition by Geographical Region



Abbreviations: MSK = Musculoskeletal

Figure 20. Incidence of Fit Notes Due to a MH Condition by Geographical Region



Abbreviations: MH = Mental Health

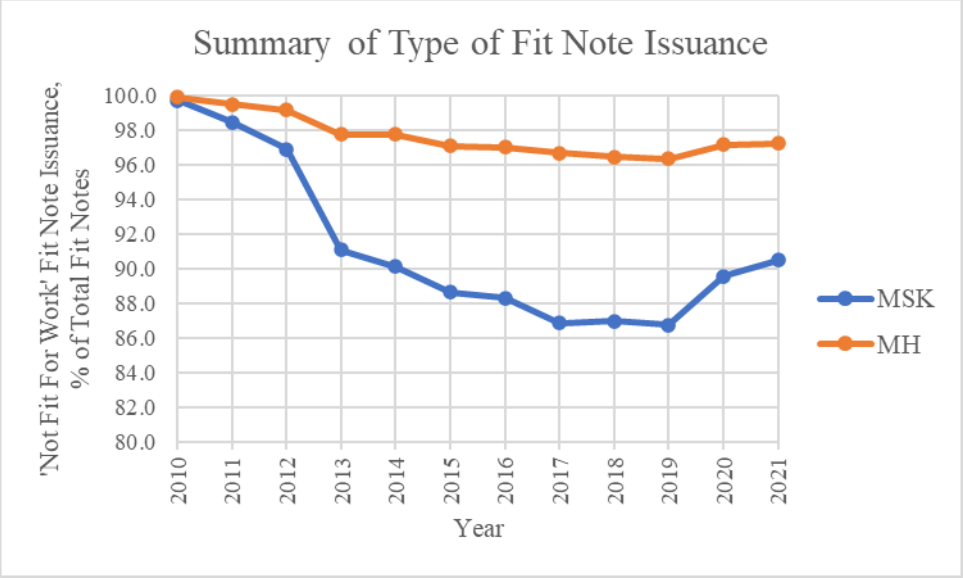
For MSK conditions, no clear trends in incidence by geographical region of GP practice were discernible, especially in later years where the region-specific incidence rates converged.

In contrast, for MH conditions, geographical differences were more apparent. In particular, London consistently had the lowest fit note incidence, whilst the Northwest and Northeast generally had the highest incidence.

6.3.6 Type of Incident Fit Note Issuance

Nearly all incident fit notes were issued as not fit for work in 2010 and 2011, as shown in Figure 21. However, whilst nearly all incident fit notes for a MH condition (>96.4% per year) remained as not fit for work up to 2021, there was an increase in the proportion of fit notes issued as 'maybe fit for work' (peaking at 13.2% in 2019) for a MSK condition.

Figure 21. Summary of Type of Incident Fit Note ('Not Fit' versus 'Maybe Fit' for Work) Due to MSK and MH Conditions Between 2010 and 2021



Abbreviations: MH = Mental Health; MSK = Musculoskeletal

6.4 Discussion

6.4.1 Summary of Findings and Comparison with Other Research

In this study incidence of fit note issuance due to a MSK or MH condition was investigated over 2010-2021 for an English population, using the CPRD Aurum dataset.

The key findings were that incidence of fit notes due to a MH condition were steadily increasing from 2014 to 2021, whilst those due to a MSK condition were decreasing. No clear differences in fit note incidence due to a MSK condition were observed by sociodemographic characteristics. Whereas incidence for MH conditions was shown to be higher in younger age groups, females, and in the Northeast and Northwest regions of England. In particular, the 16-25 years female subgroup was identified as having the highest fit note incidence due to a MH condition. London exhibited the lowest incidence of fit notes due to a MH condition.

The interpretation of these findings is now presented and has been informed by an investigation of sociological literature concerning work.

One possible explanation is that the increase in fit note incidence due to MH conditions may have been affected by austerity in the UK. Austerity is defined as actions and policies implemented by the state to reduce "spending on public expenditure with the precise aim of reducing governmental budget deficit".[184(p2)] Hall (2019) posited that austerity in the UK was "inextricable from the Global Financial Crisis (GFC) and the period from 2010 onwards",[184(p4)] which corresponds to the timeframe of this study.

Austerity can have far-reaching financial consequences, including increased unemployment and debt levels, as well as reductions in income. These financial issues

can result in greater MH problems, through reduced well-being and resilience, as well as increased MH needs.[185,186]

The observed increase in incidence of fit notes due to a MH condition may also be attributed to an increase in MH awareness campaigns in society in general, which promote reduction of stigma around MH. For example, the annually commemorated World Mental Health Day is still relatively new, having inaugurated on 10th October 1992. A 2015 House of Commons briefing also stated that although improvements were still needed, "anti-stigma campaigns and the growing profile of mental health issues in recent years appear to have gone some way to changing views and dispelling misconceptions about mental illness".[187(p48)]

In contrast, the reduction of fit note incidence due to MSK conditions may be explained by the more longstanding knowledge of management of MSK conditions as compared to MH conditions. During this study's timeframe NICE published new guidelines for managing low back pain and sciatica in 2016.[188] These guidelines included a chapter on the effectiveness of different return to work programmes, and ultimately, NICE recommended that a return to work should be facilitated and promoted as a non-invasive treatment for low back pain and sciatica.[188] However, it is unclear whether these guidelines may have led to a reduction in fit note incidence due to a MSK condition.

Conversely, there is evidence that MH conditions can be harder to manage. For example, Gabbay et al (2020) posited that there is "a perceived shortage of skills in the diagnosis and treatment of depression in primary care in the UK",[189(p662)] and that it can be difficult for a HCP to know how a person's working environment interacts with their mental health and the best way to approach this.

An abrupt single year decrease for 2020 overall incidence rates was also observed in this study, likely due to the effect of the COVID-19 pandemic. During this time, online 'isolation notes' could be used in lieu of a fit note if a person had symptoms of COVID-19.[190] Furthermore, a furlough scheme was implemented in 2020, whereby employers who could not provide a working role for their employees due to COVID-19 restrictions could instead 'furlough' their employees, whereby the UK Government paid 80% of the employees wage whilst they could not work.[191] This temporary reduction in the active workforce is likely to have led to less fit notes being issued. Additionally, for those that were able to work and were given the option to work from home in 2020, this flexibility in working may have led to more people self-managing their sickness absences, which could have further contributed to a reduction in incident fit notes.

In terms of sex, females were shown to have a higher incidence of fit notes due to a MH condition. Pattyn et al (2015) showed in their study, that there was a gender gap in use of MH services, whereby males sought less help as they had a negative attitude towards the value of psychotherapy.[192] Males associated help-seeking behaviour with femininity and shame, and rated self-management as the more useful treatment option for other males.[192] Moreover, Pattyn et al found that females contributed to the preservation of these masculine tendencies, by deeming self-management to be a better treatment option for males, but psychotherapy as more useful to other females.[192]

Then, in terms of stratifications of both age and sex, a young female subgroup was identified in this study as the highest risk group for incident fit notes due to a MH condition. This was also seen in a NHS Digital report published in 2016,[193] of the Adult Psychiatric Morbidity Survey 2014, which identified young females as a high risk group in terms of higher rates of: common mental disorder, self-harm, bipolar disorder, and

positive screening for post-traumatic stress disorder. This report also found that the gap between young females and young males had increased over time.

Additionally, it was also shown in this study that for incident fit notes due to either a MSK or MH condition, the oldest group (aged 56-66 years), consistently had the lowest incidence. The healthy worker effect may explain this.[194] If a person had reached the age of 56 years without ever having been issued a fit note, it is likely that they would be a rather healthy individual, and therefore it would be plausible that they would reach 66 years without being issued a fit note.

The final incidence stratification, geographical region, showed that fit note incidence due to a MH condition was highest in Northeast and Northwest regions of England, and lowest in London. This was suggestive of a possible North-South divide of incidence, which has similarities with the work of Parker et al.[21] This study, mentioned previously in Section 1.2.3, showed that HWLE was higher in Southern parts of England, and lower in Northern parts (particularly in the Northeast of England). This could be attributable to London and the Southeast of England leading the country's growth, and being the regions with the highest income and productivity,[195] whilst the Northeast for example, had the lowest economic competitiveness of all regions of England in 2021.[196]

These regional differences may have also been impacted by unequal austerity measures. For example, the Office of National Statistics data showed that there were more than 500,000 jobs lost in the public sector from June 2010 to September 2012, and more than 35% of these job losses occurred in the North of England.[197]

The latest fit note summary data from NHS Digital published in January 2024,[38] also concurred with the geographical region incidence results of this study. The Northwest of

England had the highest average rate of fit note issuance over Quarter 2, 2023-2024, with 3,162 fit notes issued per 100,000 persons. Whilst London had the lowest regional fit note issuance rate over this time, with 1,705 fit notes issued on average per 100,000 persons.

Finally, this study also showed that almost all incident fit notes in 2010 and 2011 were issued as 'not fit' rather than 'maybe fit for work' fit notes. This is likely to have been due to the fact that the new fit note, which for the first time offered a 'maybe fit for work' option, was only introduced on 6th April 2010. Hence, perhaps not enough time had elapsed for HCPs to be confident in prescribing this new 'maybe fit for work' option.

However, it was also shown that for MH conditions, >96.4% of yearly incident fit notes were issued as 'not fit for work'. Whilst for MSK conditions, 'not fit for work' fit notes were still the most common type of incident fit note, but there was an increasing amount issued as 'maybe fit for work' over time. Thus, there seemed to be a comparatively greater uptake for HCPs to prescribe this 'maybe fit for work' fit note for MSK conditions, as opposed to for MH conditions.

Low uptake of the 'maybe fit for work' option is demonstrated in the latest NHS Digital fit note summary data report,[38] with 94.1% of fit notes issued as 'not fit for work' from April 2021 to September 2023. Similarly, Gabbay et al (2020) also agreed that the current uptake of the 'maybe fit for work' option was rare.[189] Gabbay et al went on to say, in reference to the 'maybe fit for work' option: "when such advice is given, it tends to be more prevalent when physical health problems are the reason for the certification of sickness absence",[189(p663)] which further concurs with the findings of this study.

6.4.2 Study Strengths and Limitations

The study had the benefit of using a large sample (n=375,113 incident fit notes due to a MSK or MH condition were analysed in total), from a data source that has been shown to be representative of the general English population.[128] Therefore the findings are expected to be generalisable to England as a whole. Furthermore, this study had the strength of being one of the few studies to investigate patterns of fit note issuance in an English population, and the first to do so over a recent time frame.

However, despite the large sample size, there was a discrepancy observed in which there were considerably less individuals with an incident fit note due to a MSK or MH condition in our study than expected from our feasibility count. For example, the feasibility count (see Section 6.2.4) suggested that there would be approximately 130,000 and 300,000 incident fit notes issued in 2019 due to a MSK or MH condition, respectively. Yet, in the analysis there were 12,512 and 22,949 incident fit notes issued in this year due to a MSK or MH condition, respectively (as shown in Section 6.3.1).

This is posited to be mainly due to the NHS Digital estimates[183] that were used to derive feasibility counts looking at all fit notes for MSK or MH conditions, rather than incident fit notes for these conditions only. Furthermore, our study definitions of MSK and MH conditions, whilst comprehensive, excluded less common types of MSK and MH conditions (for example, MSK conditions such as elbow, shoulder and foot pain, and MH conditions such as burnout), which may further explain the discrepancy.

In addition, another contributing factor to these lower-than-expected incident fit note numbers, and one of our main study limitations, was that the reason for fit note issuance had to be inferred using medical consultation data, as such data was not available

directly from the fit note in CPRD Aurum. A fit note was assumed to be due to a MSK or MH condition if there was a MSK or MH consultation, respectively, in the two weeks prior to or on the first ever fit note date.

In a study by Lewis et al (2015),[198] the authors faced a similar issue in not having direct access to reason for fit note data. Lewis et al (2015)[198] investigated association of sickness certificate issuance for low back pain patients against clinical and cost-related outcomes. Persons who agreed to participate in the study were sent a baseline questionnaire (that contained the clinical and cost-related outcomes) to complete in the week following their low back pain consultation, and this was then linked to medical records to identify sickness certification. To allow for a delay of up to one week in baseline questionnaire issuance from low back pain consultation, as well as the mailing delay in the return of the questionnaire, the study authors opted for a linkage approach of searching for sickness certification records in the 31 days prior to the questionnaire being received.

Whilst it can be problematic deciding how best to link medical consultation and sickness absence data, Dorrington et al (2021)[78] argued that taking reason for absence data directly from a fit note can also be also limiting too, as many fit notes do not provide sufficient information for a sickness absence reason to be properly coded into an EHR database in the first instance.[78] Also, the information on the fit note might not always depict the true reason for sickness absence, for example if a patient is sensitive to information about a MH problem being shared with their employer the HCP might write another reason on the fit note. Furthermore, multimorbidity is often not recorded on a fit note, albeit the fit note itself does not restrict this. Therefore, using medical consultation history can be a better way to conceivably extract a more accurate reason

for the fit note, and to allow for multimorbidity information too. Albeit in this study a low count of fit notes due to a MSK and MH condition simultaneously was observed.

Another limitation was that a fixed age cut-off of 56-66 years was used to define the oldest age cohorts from 2010 to 2021 in this study, to keep the study population for working age only. However, this is not strictly correct with respect to SPA changes from 2010 onwards. A complex SPA roll-out system was used in practice, based on date of birth, whereby the female SPA changed from 60 to 66 years over two-monthly increments. Hence there might have been some females included in this analysis that had retired and were no longer economically active, thus exacerbating the reported healthy worker effect. However, this is not likely to have greatly impacted upon the reported results.

Further to this point, it is possible that the denominator population of this study included some individuals that might not have been economically active for other reasons too, such as leaving the workforce.

There was also a limitation concerning the definition of incidence. In this study incidence was based on individuals that had never been issued a fit note before, rather than individuals that had never been issued a fit note for a MSK or MH condition. If the latter approach had instead been taken, different patterns in incidence rates might have been observed. However, using such a broad incidence definition would have also permitted individuals who received multiple previous fit notes, perhaps for a chronic illness (due to a reason other than a MSK or MH condition), to be treated the same as individuals who received no previous fit notes, at the point at which a first MSK or MH condition fit note was issued. Having such wholly different participant populations

mixed together would have made inferences from these study results more challenging, and comparisons between the MSK and MH condition results more difficult.

Furthermore, other covariates could have potentially been explored when stratifying incidence rates, instead of only age, sex, and geographic region. For example, in Chapter 8, the same three covariates as in this study were used to assess for presence of characteristic associations with trajectories of work absence, yet further covariates were also considered (covering health characteristics, types of treatment received and comorbidity). However, to include further covariates was problematic in this study, because the denominator in the incidence rate calculation for a particular year, is the total number of English practice registered patients in the CPRD Aurum database in that year. Thus, stratification of incidence rates by covariates requires the covariate data for the denominator population to be downloaded. For covariates other than age, sex, and geographic region, this is challenging due to the complexity and size of the data that would need to be downloaded. Furthermore, covariate data at a denominator population level is not always available.

6.4.3 Conclusion

In conclusion, this study has identified that there were differences in incidences of fit notes due to MSK and MH conditions from 2010 to 2021. From 2014 to 2021, fit note incidence due to MSK conditions decreased, whilst there was an increase due to MH conditions. No clear differences in fit note incidence by sociodemographic variables were observed for MSK conditions. However, for MH conditions, younger females and persons living in the Northeast and Northwest of England exhibited higher incidence, whilst London had the lowest. In particular, a 16-25 years female subgroup was identified as highest risk for fit note incidence due to a MH condition.

In the next Chapter, the main analyses of this study are conducted through derivation of trajectories of work absence. This is performed by taking a subset of individuals from this current Chapter, and following these individuals up over time to assess whether it is possible to categorise them into subgroups based on their fit note issuance patterns.

Chapter 7. Study 2: Deriving Optimal Trajectories of Work Absence

7.1 Study Aim and Objectives

In this Chapter, through a retrospective cohort study, one of the main aims of this thesis was addressed: to derive, and compare using different statistical methods, common longitudinal trajectories of work absence as measured by receipt of fit notes, for a population consulting their HCP with a MSK or MH condition.

The overall aim of this Chapter was to determine whether it was possible to derive work absence trajectories due to a MSK or MH condition, then the specific objectives were threefold:

1) To determine the most appropriate time intervals for deriving trajectories of work absence

2) To determine the most appropriate statistical method for deriving trajectories of work absence

3) To assess whether the derived absence trajectories were a function of the reason for absence onset (i.e. whether the trajectories based on incident absence due to a MSK condition differed from those due to a MH condition)

7.2 Methods

7.2.1 Defining the Study Population

Trajectories of absence were derived based on the same baseline population of individuals with a first ever fit note due to a MSK or MH condition from study 1 (i.e.,

the same inclusion/exclusion criteria were used as in Section 6.2.2). Although, in this Chapter, further consideration towards missing data was applied than in the previous study, as explained later in Section 7.2.8.

An overall follow-up period of up to three years was chosen when deriving trajectories of absence. This was deemed a suitable period to explore longer-term patterns of work absence of more than one year as well as shorter-term absences (choice of trajectory follow-up length is discussed further in Section 7.2.4).

However, rather than taking all individuals with a first ever fit note issued from 2010 onwards as in study 1, the years were restricted in this study from 2016 to 2018 (individuals with a first ever fit note during this time period were pooled into a single cohort). This was done to reduce the amount of data being used in the trajectory analysis models (as such analysis is computationally intensive), and to ensure that the three-year follow-up data was as recent as possible.

For example, if there was a participant from study 1 with a first ever fit note (due to a MSK or MH condition) issued on 31st December 2018, they would be included in this study, and their patterns of fit note issuance (for any reason) then analysed up to 31st December 2021.

7.2.2 Sample Size Considerations

Initially, up to 130,000 and 300,000 individuals per year were expected to be issued a first ever fit note due to a MSK or MH condition, respectively (as described in Section 6.2.4).

Therefore, to reduce the computational intensity of the trajectory analysis, it was originally planned to randomly select 50,000 eligible individuals who were issued a first

fit note due to a MSK condition during 2016 to 2018, and 50,000 for a MH condition. However, as explained in the Discussion Section of the previous Chapter, the observed numbers of individuals receiving an incident fit note were substantially lower than originally estimated (see Section 6.4.2). Hence, in this study, it was not necessary to take a random sample of eligible individuals. Rather, all eligible individuals were utilised (this change is reflected in the amended CPRD data request protocol in Appendix F, as mentioned in Section 4.3).

The subset of individuals from study 1, that had a first ever fit note due to a MSK or MH condition from 2016 to 2018 are described in Table 19. Thus, there were n=42,264 baseline individuals available for use when deriving trajectories of absence due to a MSK condition, and n=60,567 due to a MH condition (the final numbers used are explained later in Section 7.2.8).

Table 19. Number of Individuals Available for Trajectory Derivation Analyses

Year	First Ever Fit Note Due to a MSK Condition, n	First Ever Fit Note Due to a MH Condition, n
2016	15,100	19,442
2017	13,942	20,302
2018	13,222	20,823
Total	**42,264**	**60,567**

To demonstrate the statistical power present in this study, examples of the precision of trajectory prevalence (based on 50,000 baseline individuals) are provided, with 95% CIs estimated using a normal approximation to the binomial calculation:

- 10% trajectory prevalence (9.74%, 10.26%)
- 30% trajectory prevalence (29.60%, 30.40%)

- 50% trajectory prevalence (49.56%, 50.44%)

Thus, these narrow 95% CIs suggest trajectory prevalence was estimated with a high level of precision in this study.

7.2.3 Exploration of a Continuous Fit Note Definition

Initially, it was planned for trajectories of work absence to be derived based on both a continuous definition (cumulative days of absence over follow-up, or average days of absence in a given time interval), as well as a binary definition (yes/no for fit note issuance in a given time interval). A final optimal model would then be chosen, taking into consideration all trajectory models fitted for either of these two types of fit note definition.

However, it was ultimately not possible to use a continuous fit note definition in this thesis, despite attempts to generate continuous fit note data, as explained in this Section.

Generating a continuous work absence definition for trajectory derivation requires fit note duration. Though, as mentioned in Section 6.2.7, fit note duration was largely missing in the available CPRD data for this thesis (only 0.43% of baseline individuals with an incident fit note from 2010-2021 due to a MSK and/or MH condition had a reported fit note duration). Thus, to assess the feasibility of modelling trajectories based on a continuous repeated measure of absence, an approach to estimating missing fit note duration was explored.

For this exploration, a subset of data from study 1 was used, for individuals who had an index fit note due to a MSK condition in 2016 (n=15,100 individuals). These individuals were issued a total of n=53,773 fit notes during a three-year follow-up (this

includes the index fit note due to a MSK condition, as well as subsequent fit notes issued for any reason during follow-up).

To differentiate between multiple absence episodes for the same participant, a rule was applied of there being at least a six-month gap between fit notes for it to be considered as a new episode.

A four-step hierarchical process was then used to generate a fit note duration (in days):

Step 1) Where present, CPRD fit note duration in days was used.

Step 2) Otherwise, for a particular participant, if there was more than one fit note in a given absence episode, the start date of the latter fit note was used to infer fit note duration of the former. An assumption was made that the end date of the first fit note was equal to the start date of the second. This process was then repeated between all successive fit notes in a given absence episode (second and third fit notes, third and fourth, etc), until the final fit note in a given sickness absence episode was reached (at which point there was no successive fit note start date available for use).

Step 3) If the last fit note in a given sickness absence episode was reached for a particular participant, available median fit note duration data (from steps 1 and 2) was matched from individuals based on the same absence episode-fit note number. Ideally, this matching would also incorporate stratification by sociodemographic characteristics (sex, age, and region) if there was sufficient data.

Step 4) Finally, if there was insufficient median fit note duration data to allow step 3 to be performed, a standard value of a 14-day fit note duration was applied.

After performing step 1, n=6 fit note durations were retrieved from the CPRD fit note duration data that was present for these n=53,773 fit notes (i.e., fit note duration was non-missing for only 0.01% of this 2016 MSK cohort).

Next, after performing step 2, and assuming the start date of a successive fit note within the same absence episode was the end date of the former fit note, n=32,043 fit note durations were computed. This left n=21,724 fit note durations to be estimated (40.4% of the data) with steps 3 and 4.

Then, in preparation for performing step 3, the n=53,773 fit notes issued over the 3-year follow-up for the cohort with an index fit note due to a MSK condition in 2016 were first summarised by absence episode number and fit note number (Appendix I).

Individuals in this 2016 MSK condition cohort experienced up to five absence episodes, and around 94% of their three-year fit note data was contained in the first two absence episodes (Appendix Figure I.1). Each absence episode contained up to 36 fit notes, with approximately 90% of the data contained in the first 7 fit notes (Appendix Figure I.2).

However, counts within combinations of absence episode and fit note number became small (results omitted for brevity), even more so when further stratified by age, sex and region. Hence it was concluded that step 3 was too unrealistic to conduct.

Thus, at this stage, through discussion between AL and his supervisory team, it was decided that using a continuous definition of fit notes for trajectory derivation with the available CPRD Aurum data was not feasible.

This decision was also made concerning step 2 of the approach. Although this step did allow a large amount of missing fit note duration data to be generated (60%), the underlying assumption here was that an absence episode was six months in duration and therefore there had to be a gap of at least six months in recorded fit notes for one episode to end. Such a lengthy duration may be a strong assumption, as fit notes that are issued months apart may in fact relate to different episodes. Conversely, if the definition of the length of an absence episode duration was reduced, this would have resulted in more of the n=53,773 fit note durations being computed by steps 3 and 4 in this dataset, instead of step 2, which may then pose new problems (as the step 3 median fit note durations largely originate from step 2).

Also, step 4 is the least favourable option, as all individuals are assigned the same fit note duration value, which eliminates all heterogeneity in fit note duration, and rather hinders the idea behind looking at trajectories based on duration. Furthermore, latest data from NHS Digital (from April 2021 – September 2023) showed that for all fit notes issued due to a MSK or MH condition, approximately 50% of such fit notes have duration of more than one month.[38] This further suggests that using a 14-day proposal is not plausible.

Therefore, using the described approach to generate fit note durations, for the process to have worked better, more fit note durations from steps 1 and 2 were required. This would have meant that the inferences in step 3 were based on more data and also that both steps 3 and 4 were used less.

To conclude, it was decided that the best course of action with the data available, was to reject basing trajectory derivation on a continuous fit note definition. Instead, a binary fit note definition was used in this study, utilising data on number of fit notes issued.

Application of this binary fit note definition in trajectory derivation is explained in the next Section.

7.2.4 Defining Trajectory Follow-Up and Interval Lengths

In this study, trajectories were defined beginning with an incident fit note issued due to either a MSK or MH condition, and then following individuals up over time to assess their patterns of fit note issuance (due to any reason).

The repeated fit note measure used to assess patterns of absence was a binary yes/no for fit note issuance (coded as 1/0) in a given time interval during follow-up. The initial MSK or MH condition incident fit note was excluded from the trajectory definition, as all individuals received this.

Two different follow-up lengths of either one- or three-years post index date were used, and with fit note issuance then assessed in these follow-up periods based on two-, three-, or six-monthly recurring intervals, as explained in this Section.

Generally, when performing trajectory derivation, there is a balance required between choosing too few time intervals in a given follow-up period, compared to too many intervals.[173] If there are too few intervals, more sophisticated patterns (of work absence over time, in the context of this study) might be missed. Conversely, having too many intervals might result in some intervals containing sparse data, which affects the accuracy of trajectory derivation. For example, in the work by Strauss et al,[199] the study authors chose to derive trajectories of multimorbidity based on a three-year follow-up and using recurring six-monthly intervals.

It is often useful to test the effect of different follow-up lengths and corresponding time intervals for best operational performance. To inform the choice of appropriate

trajectory follow-up lengths and time intervals to test in this study, data exploration was first performed using a subset of the MSK condition incident absence cohort from study 1.

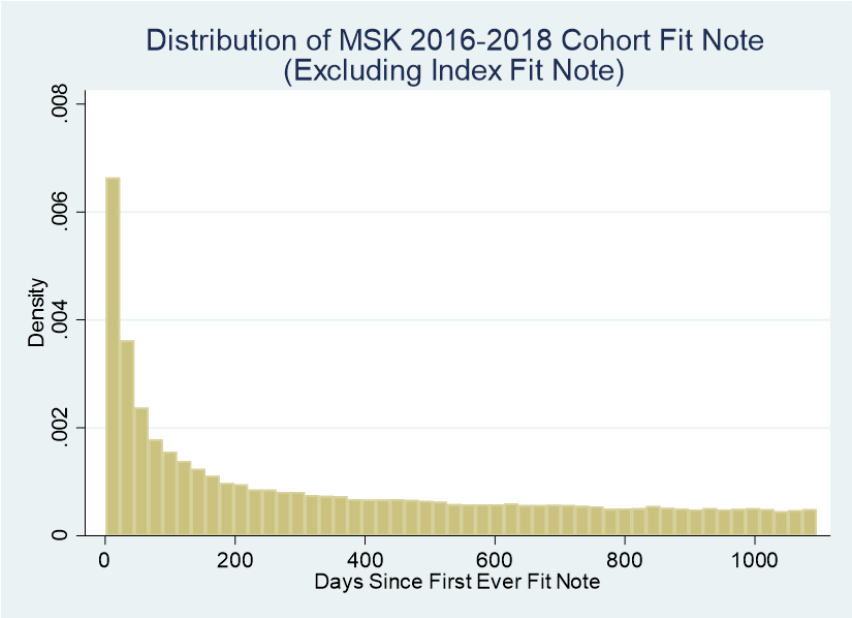
Using the MSK 2016-2018 cohort for exploration (i.e., individuals who received a first ever fit note due to a MSK condition from 1st January 2016 to 31st December 2018), there were n=42,264 individuals and they received a total of n=103,825 fit notes over a three-year follow-up (excluding the index MSK fit note). Fit note issuance over time is summarised in Table 20, and the distribution in Figure 22.

As the majority of fit notes were issued during the first year of follow-up (59%) and less in years two and three, it was decided to test the effect of using two different follow-up periods in this thesis. Firstly, a short-term follow-up of the first twelve months from index date, and secondly, a longer-term follow-up of three years from index date.

Table 20. Summary of MSK 2016-2018 Cohort Fit Notes Over Three Year Follow-Up (Excluding Index Fit Note)

Time Since Index Fit Note, Months	n (%)	Cumulative %
1	18,888 (18.19%)	18.19%
2 to 3	14,514 (13.98%)	32.17%
4 to 6	12,053 (11.61%)	43.78%
7 to 12	15,821 (15.24%)	59.02%
13 to 18	12,502 (12.04%)	71.06%
19 to 24	10,944 (10.54%)	81.60%
25 to 30	9,909 (9.54%)	91.14%
31 to 36	9,194 (8.86%)	100%
Total	**103,825 (100%)**	

Figure 22. Histogram of MSK 2016-2018 Cohort Fit Notes Over Three Year Follow-Up (Excluding Index Fit Note)



Then, for these two different follow-up lengths, the next step was to define appropriate time intervals. The chosen time intervals to be tested, taking into account the distribution of the observed fit note issuance data, were fivefold as shown below in Table 21.

Table 21. The Five Interval Approaches Used for Trajectory Derivation in this Study

Interval Approach	Year One Follow-Up Length	Years Two-Three Follow-Up Length
1 (Short-Term)	Three-Monthly	-
2 (Short-Term)	Two-Monthly	-
3 (Long-Term)	Six-Monthly	Six-Monthly
4 (Long-Term)	Three-Monthly	Six-Monthly
5 (Long-Term)	Two-Monthly	Six-Monthly

Two different interval approaches (1 and 2) were tested for a one-year follow-up (either two- or three-monthly recurring intervals). Monthly-recurring intervals were also considered, but rejected, as latest NHS Digital showed that 50% of fit notes issued due

to a MSK or MH condition have duration > one month, hence absence could be underestimated with these intervals.

Then for the three-year follow-up, three different interval approaches were tested (approaches 3 to 5), using two-, three- and six-monthly interval combinations. In particular, in interval approaches 4 and 5, the three-year follow-up was partitioned such that shorter two- or three-monthly intervals were tested for the first year of follow-up and longer six-monthly intervals for years two and three.

In conclusion, trajectory models were fitted using a binary definition of fit note issuance (for any reason) based on each of these five approaches to time intervals. In the next Section, the process of using these five interval approaches to reach an optimal trajectory model choice is described.

7.2.5 Model Building Strategy

The general trajectory model building strategy used, was as recommended by Wickrama et al,[141] to start from the simplest type of model, and progressively increase the complexity:

- Step 1: Identify an appropriate LGCM
- Step 2: Perform LCGA
- Step 3: Perform GMM

A flowchart of the model building strategy used to determine the optimal LGCM and/or LCGA (i.e. steps 1 and 2 from above) is shown in Figure 23. This strategy was applied first for individuals with an index MSK fit note from 2016 to 2018, and repeated separately for each of the five chosen approaches to time intervals from Section 7.2.4.

Then this process was also repeated for individuals with a first ever fit note due to a MH condition from 2016 to 2018.

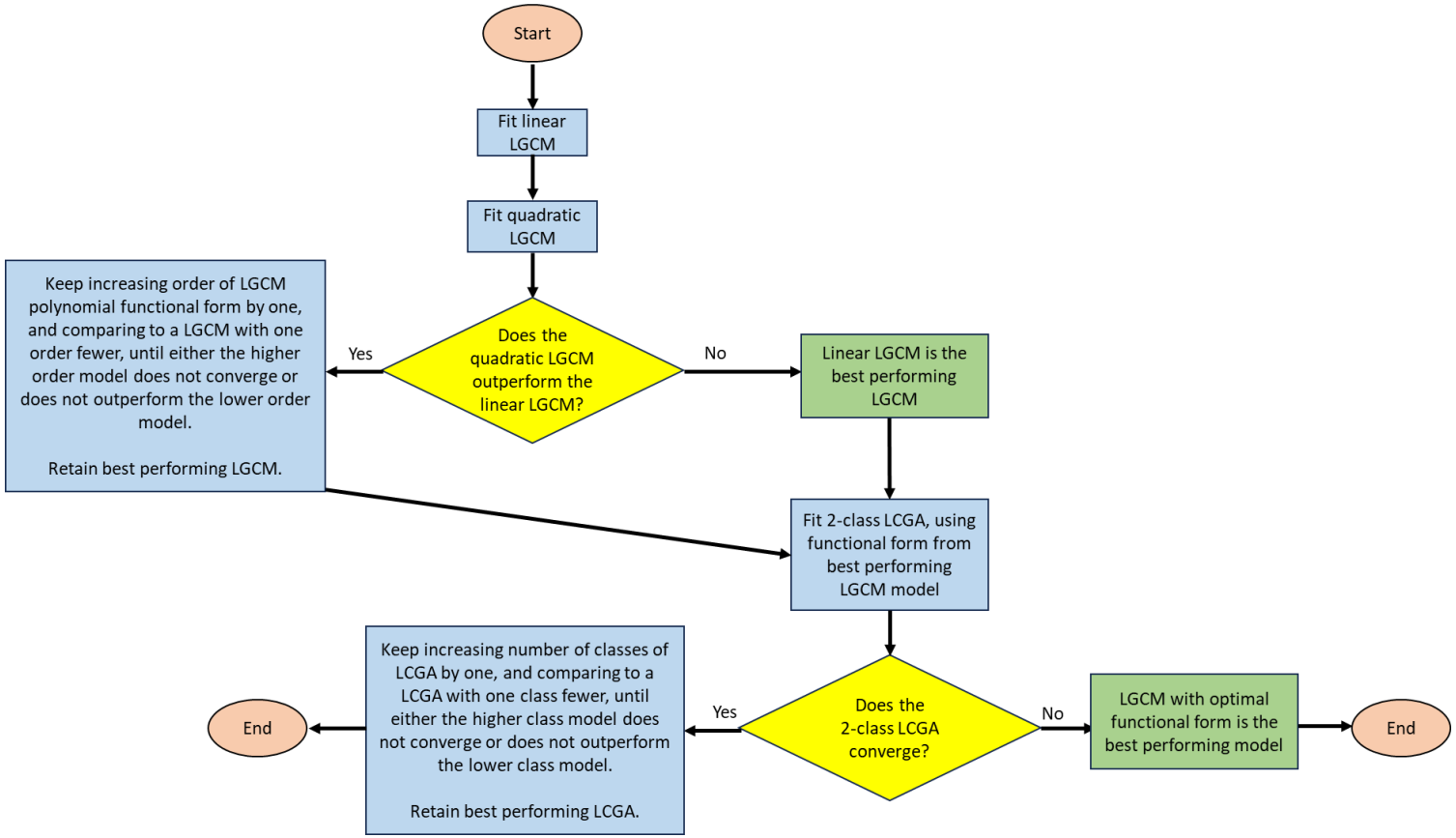
Now, the model building strategy is explained in further detail.

An initial LGCM was fitted by assuming a linear functional form (with model equations and path diagram as described earlier in the trajectory methods Chapter in Section 5.1.4).

Then, the polynomial functional form of the LGCM was increased to the next order (i.e., a quadratic form was now assumed, as shown earlier in Figure 7 and Box 3). If this model with quadratic functional form converged, its performance was contrasted against that of the LGCM with linear functional form, and the model deemed the best fitting of the two retained.

The criteria used to assess LGCM quality were as detailed in the trajectory methods Chapter, in Table 8. Briefly, better quality LGCMs were those that fitted well graphically, and had lower RMSEA and SRMR values, as well as higher CFI and TLI (see Section 5.1.6 for a more detailed explanation of these criteria and the cut-offs used).

Figure 23. Model Building Strategy Flowchart Used to Determine Optimal LGCM or LCGA Model



Abbreviations: LGCM = Latent Growth Curve Model; LCGA = Latent Class Growth Analysis.

This process was then repeated, each time building up the complexity of the functional form of the LGCM by increasing the order of the polynomial by 1 and comparing successive LGCMs, until increasing the complexity either did not improve the original model or the higher order model did not converge.

Next, a LCGA model was fitted, using the same functional form as in the final chosen LGCM model. Initially, the simplest LCGA model was fitted, the two-class model (for reference, the model equations and path diagram for a linear LCGA model with *C* classes were shown earlier in Figure 9 and Box 4, respectively, in Section 5.2.1).

If the two class LCGA model converged, then a three class LCGA was fitted. The number of classes in the LCGA model were progressively increased, each time comparing successive models, until either convergence was not possible, or the higher class LCGA model was deemed to perform worse than the lower-class model.

The criteria used to evaluate competing LCGA models, were as stated earlier in Table 9 (from Chapter 5 also). Briefly, better performing LCGA models were those that had lower AIC and BIC values, statistically significant LRT tests, entropy and average posterior probabilities ≥0.7, amongst other criteria (see Sections 5.2.5 and 5.2.6 for more details).

However, if the two-class LCGA did not converge, and after best possible efforts to overcome non-convergence (such as increasing the number of random sets of starting values and number of iterations, as well as additional modelling specifically for approaches 4 and 5, as explained in the next Section), the LGCM with the optimal functional form was deemed the final chosen model in this analysis.

7.2.6 Additional Modelling for Approaches 4 and 5

Before finalising the choice of an optimal LCGA model with the strategy discussed thus far, an alternative form of the LCGA model was also considered specifically for interval approaches 4 and 5 from Table 21. As these two interval approaches involved a combination of shorter-term (recurring two- or three-monthly) and longer-termed (recurring six-monthly) intervals in the same LCGA model, if there were any model convergence issues or evidence of particularly poor model fit, a piecewise model was considered (see earlier Section 5.1.5 for more details on this type of model).

Given that there was a divide with most of the trajectory derivation fit note issuance data concentrated in the first year of follow-up, and less in years two and three (as shown for the MSK 2016-2018 cohort in Section 7.2.4), in any piecewise models used for approaches 4 or 5, two ‘pieces’ were created in the time continuum, using a pivotal point of 12 months. In other words, one ‘piece’ of the model was based on the first year of follow-up, and the other on years two and three.

The same final functional form decided from the LGCM was retained (as with the conventional LCGA models discussed thus far too), but now different latent slopes were permitted for each of these two time periods, within one overall (piecewise) LCGA model.

If a piecewise LCGA was performed, for completeness, a piecewise LGCM was also fitted to the data. An example piecewise model for a LGCM based on two linear slopes and interval approach 4 was shown earlier in Figure 8, in Section 5.1.5.

To conclude the overall model building strategy up to the end of step 2, a final optimal class LCGA model was then chosen, through consideration of the performance of all

LCGA class models derived across all five time interval approaches (using performance criteria from Table 9). This process was repeated separately based on LCGA models for an index MSK condition fit note from 2016 to 2018, and an index MH condition fit note.

7.2.7 GMM

Finally, in the third step of the model building process, once the optimal LCGA class model was chosen from the above steps, an equivalent GMM class model was also applied.

Application of GMM involved repeating the same optimal LCGA model, but now allowing there to be variation within a trajectory class, rather than assuming all individuals in a class followed the same pattern (see Section 5.2.2 for further details on GMM methodology).

However, due to the increased complexity and computational intensity of fitting a GMM model,[61] this step was performed as an exploratory process. The same guidelines for assessing LCGAs from Table 9 were also used to assess the GMM model (if it converged). The GMM model was compared against the optimal class LCGA model, and the better performing of these two models was then the final optimal model in this study.

Firstly, the simpler version of the GMM model was attempted, a GMM-CI model (as also described earlier in Section 5.2.2). If this model outperformed the optimal class LCGA model, then the more complex form of GMM-CV model was attempted next (and compared to the GMM-CI model, with the best performing model retained as the optimal model).

For the two final optimal trajectory models, for incident fit notes due to either a MSK or MH condition, a summary of the observed patterns of fit note issuance for individuals within a trajectory class were also provided, to assess individual variability of the final trajectory classes. Furthermore, if the optimal model was based on an interval approach that used year one follow-up data only, the observed proportions of fit note issuance for these trajectory classes in years two and three of follow-up were summarised.

7.2.8 Missing Data

In the previous study on incidence rates of absence from 2010 to 2021, individuals who were aged ≥67 years old, died, de-registered from their GP practice, or had a last collection date during or before the year of the incidence rate calculation were excluded from the study (see Table 14 in Section 6.2.2).

However, an alternative approach could have been used, whereby such individuals could have been included if they had some time at risk during the follow-up period. For example, in the 2010 incidence rates calculation, if an individual died halfway through 2010, rather than exclude them completely, they could have been included for the first six months of follow-up and then censored after this point.

Nonetheless, censoring was not deemed of concern, given that the magnitude of all calculated incidence rates was low. This was due to the sizeable discrepancy between the numerator and denominator counts (the denominator contained at least 6.5 million individuals in each yearly calculation, whereas the numerator contained a maximum of 25,000).

In contrast, in this study, the sample size is much smaller, as this study is based solely on following-up the numerator cohort from the incidence rates study. Therefore, even

small amounts of missing data could potentially have a large impact on the quality of this study's results.

Thus, the approach taken to the main analyses in this study was to include all available data, including from individuals who did not have complete follow-up. If a participant incurred any of the four exclusion reasons mentioned above after entering the cohort (relating to age ≥67 years, death, de-registration, or last collection date), their data was censored from the time interval in which the exclusion reason first occurred, as well as all succeeding time intervals during the follow-up.

To affirm the integrity of this approach, the results from this study's main analyses were compared for similarity against those from two different sensitivity analyses. The first sensitivity analysis involved re-running all of the trajectory derivation analyses excluding individuals who incurred an exclusion reason during the calendar year of their index fit note, and assuming everyone included had a complete three years of follow-up.

Secondly, a stricter sensitivity analysis was also conducted, whereby individuals were excluded if they incurred any exclusion reason at any point during follow-up (i.e. a complete case analysis), not just during the calendar year of the index fit note. This second sensitivity analysis was only conducted for the chosen optimal trajectory models (separately for the cohorts with an index fit note due to a MSK or MH condition).

The reporting in this study was carried out in line with the GRoLTS checklist.[173]

Input files for trajectory derivation were prepared using Stata MP version 17.0, and all trajectory derivation analyses performed using Mplus Version 8.9. The trajectory derivation models were fitted using Mplus default estimation options.

The study results are presented in the next three Sections as follows:

- In Section 7.3, the trajectory model results based on an incident MSK condition fit note are described, separately for each of the five time intervals considered.
- This process is then repeated in Section 7.4 for the incident MH condition fit note cohort.
- Then, in Section 7.5, an optimal model is chosen and summarised (this Section also includes an overview of the results of the different sensitivity analyses to test alternate approaches for handling missing data).

7.3 Results: Trajectories of Work Absence Due to a MSK Condition

7.3.1 Approach 1: Three Month Intervals (Year One Only)

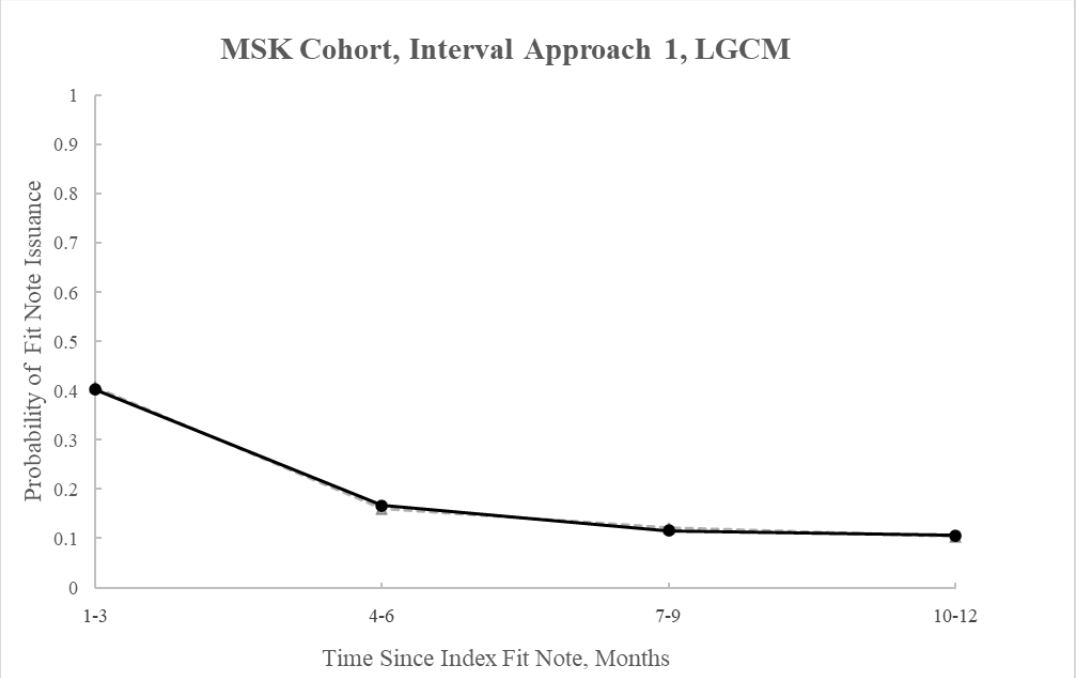
The linear LGCM fitted to year one follow-up data based on three-monthly intervals (interval approach 1), met all of the model fit indices thresholds set out in Table 8, which indicated good model fit (as shown in Table 22). Graphically, the model fitted the data well too, with no discernible difference between the observed fit note proportions and model estimated probabilities at each time interval (Figure 24). In terms of the LGCM shape, the highest probability of fit note issuance (of 0.4) occurred in the first three-month interval, this then reduced to 0.2 by month six, and further reduced steadily to end at 0.1 after the full year of follow-up.

Table 22. Model Fit Indices of Linear LGCM for MSK Cohort, Interval Approach 1 (Three Monthly Intervals, Year One Data Only)

n	42,905
RMSEA Estimate (90% CI)	0.06 (0.05, 0.07)
CFI	0.99
TLI	0.96
SRMR	0.03

Abbreviations: n = number of individuals; LGCM = Latent Growth Curve Model; MSK = Musculoskeletal; RMSEA = Root Mean Square Error of Approximation; CI = Confidence Interval; CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; SRMR = Standardised Root Mean Residual

Figure 24. Linear LGCM Trajectory Plot for MSK Cohort, Interval Approach 1 (Three Monthly Intervals, Year One Data Only)



Abbreviations: LGCM = Latent Growth Curve Model; MSK = Musculoskeletal;
The solid lines represent the probability of fit note issuance in the given time interval (model estimated data), whilst the dotted lines represent the proportion of individuals issued a fit note (observed data).

The quadratic LGCM did not run for interval approach 1, nor for any interval approach in this study (even after increasing number of random sets of starting values) – the degrees of freedom for the model were negative, and the model was not identified. This was the case for all quadratic LGCM models fitted for the MH condition index fit note cohort too. Therefore, throughout this study, the linear functional form of the LGCM was retained as the optimal form for all models. Thus, all LCGA and GMM models were also fitted with a linear functional form only.

The model fit and class meaningfulness statistics of the (linear) LCGA models for approach 1 are shown in Table 23, and corresponding plots in Figure 25. The LCGA model did not run when the number of classes was increased to four.

Both the two- and three-class LCGA models converged without issue, and had a LRT p value <0.0001 (based on both the LMR-LRT and the BLRT). This suggested that the two-class model fitted the data better than a LGCM (one-class model), and the three-class model fitted the data better than the two-class (this was further demonstrated by the lower AIC and BIC values in the three-class model). Indeed, the LRT *p* value was <0.0001 for all the LCGA models described in this study.

However, the entropy of the two- and three-class models was similar (0.64 and 0.66, respectively) and below our study guideline threshold of 0.7 (Table 9).

Table 23. Model Fit and Class Meaningfulness Statistics of Linear LCGA Models for MSK Cohort, Interval Approach 1 (Three Monthly Intervals, Year One Data Only)

	LCGA Model (n=42,905)	
	2 Class	**3 Class**
Log-likelihood	-71500	-70518
AIC	143010	141052
BIC	143053	141122
LRT[a]	p <0.0001	p <0.0001
Average Posterior Probability (Range)	0.91-0.96	0.86-0.94
Entropy	0.64	0.66
Trajectory Class Prevalences	68.9%, 31.1%	51.7%, 31.7%, 16.6%

Abbreviations: n = number of individuals; LCGA = Latent Class Growth Analysis; MSK = Musculoskeletal; AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; LRT = Likelihood Ratio Test

[a] LRT was assessed with both the Lo-Mendell-Rubin and bootstrapped methods

In the two-class LCGA, the most prevalent class identified (68.9% prevalence) contained individuals that had an initial low probability of around 0.3 of being issued a fit note in the first 3 months of follow-up, and thereafter a sustained probability of close to 0 for fit note issuance in each three-month period from months four to twelve (named ‘Single/Short Term’; Figure 25). The second trajectory class of this two-class LCGA,

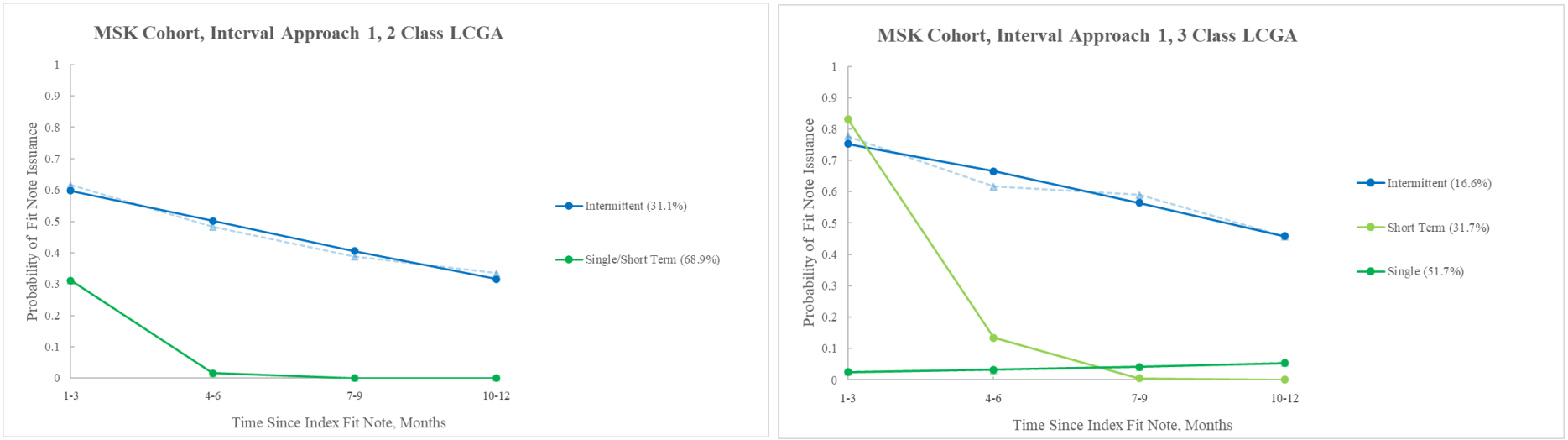
labelled ‘Intermittent’ (31.1% prevalence), was characterised by individuals with an average probability of between 0.3-0.6 of fit note issuance all throughout follow-up.

Graphically, the estimated fit note issuance probabilities for this ‘Intermittent’ class showed slight deviations from the corresponding observed proportions. In contrast, the ‘Single/Short Term’ class exhibited a near perfect graphical fit.

The three-class LCGA had a similar ‘Intermittent’ class (now with a lower prevalence of 16.6%), albeit the graphical fit showed further deterioration compared to the two-class ‘Intermittent’ (Figure 25). For example, the probability of fit note issuance during follow-up months four to six was estimated as 0.66, whereas the observed proportion was 0.62.

Furthermore, in the three-class model, it appeared that the ‘Single/Short Term’ trajectory from the two-class model had now been separated into a ‘Single’ (fit note at index date only, 51.7% prevalence) and a ‘Short Term’ class (31.7% prevalence). In the ‘Single’ class, a low probability (close to 0) of fit note issuance was sustained throughout follow-up. Whereas in the ‘Short Term’ class, there was an initial high probability of around 0.8 of receiving an additional fit note in the first three months, which then decreased to around 0.1 during months four to six, and was then sustained at almost 0 for the final six months.

Figure 25. Linear LCGA Trajectory Model Plots for MSK Cohort, Interval Approach 1 (Three Monthly Intervals, Year One Data Only)



Abbreviations: LCGA = Latent Class Growth Analysis; MSK = Musculoskeletal.
For each trajectory, the solid lines represent the probability of fit note issuance in the given time interval (model estimated data), whilst the dotted lines represent the proportion of individuals issued a fit note (observed data). Trajectory class prevalence is shown, derived from a count based on posterior probability.

7.3.2 Approach 2: Two Month Intervals (Year One Only)

The shape of the LGCM fitted for interval approach 2 was similar to the LGCM from interval approach 1, and there were again negligible visible differences between estimated probabilities and observed proportions of fit note issuance (LGCM plot shown in Appendix J, Figure J.1).

However, the LGCM fit statistics from interval approach 2 were slightly improved compared to that of approach 1 (Appendix Table J.1). In particular, the RMSEA estimate reduced from 0.06 (three-monthly intervals, approach 1) to 0.04 (two-monthly intervals, approach 2), and the TLI increased from 0.96 (approach 1) to 0.98 (approach 2). Although, both LGCMs met all of our threshold cut-off criteria for good model fit.

When LCGA was applied, models up to five-classes converged now (Table 24). The AIC and BIC values progressively decreased with each increase in class, however, the improvement was less pronounced between the four- and five-class models.

All LCGA models had similar entropy values that were greater than our guideline threshold of 0.7 (except the three-class LCGA, with an entropy of 0.67). All trajectory classes had average posterior probabilities above our threshold of 0.7 too, and each had prevalence above our threshold of 1%. The trajectories with lowest prevalences, of 2.6% and 3.7%, were from the five-class LCGA.

Table 24. Model Fit and Class Meaningfulness Statistics of Linear LCGA Models for MSK Cohort, Interval Approach 2 (Two Monthly Intervals, Year One Data Only)

	LCGA Model (n=43,130)			
	2 Class	**3 Class**	**4 Class**	**5 Class**
Log-likelihood	-90633	-88705	-87593	-87417
AIC	181276	177426	175208	174863
BIC	181320	177495	175304	174984
LRT[a]	p <0.0001	p <0.0001	p <0.0001	p <0.0001
Average Posterior Probability (Range)	0.91-0.94	0.86-0.92	0.81-0.87	0.79-0.90
Entropy	0.72	0.67	0.70	0.72
Trajectory Class Prevalences	71.3%, 28.7%	52.0%, 30.3%, 17.7%	62.1%, 23.2%, 9.5%, 5.1%	45.5%, 27.7%, 20.6%, 3.7%, 2.6%

Abbreviations: n = number of individuals; LCGA = Latent Class Growth Analysis; MSK = Musculoskeletal; AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; LRT = Likelihood Ratio Test

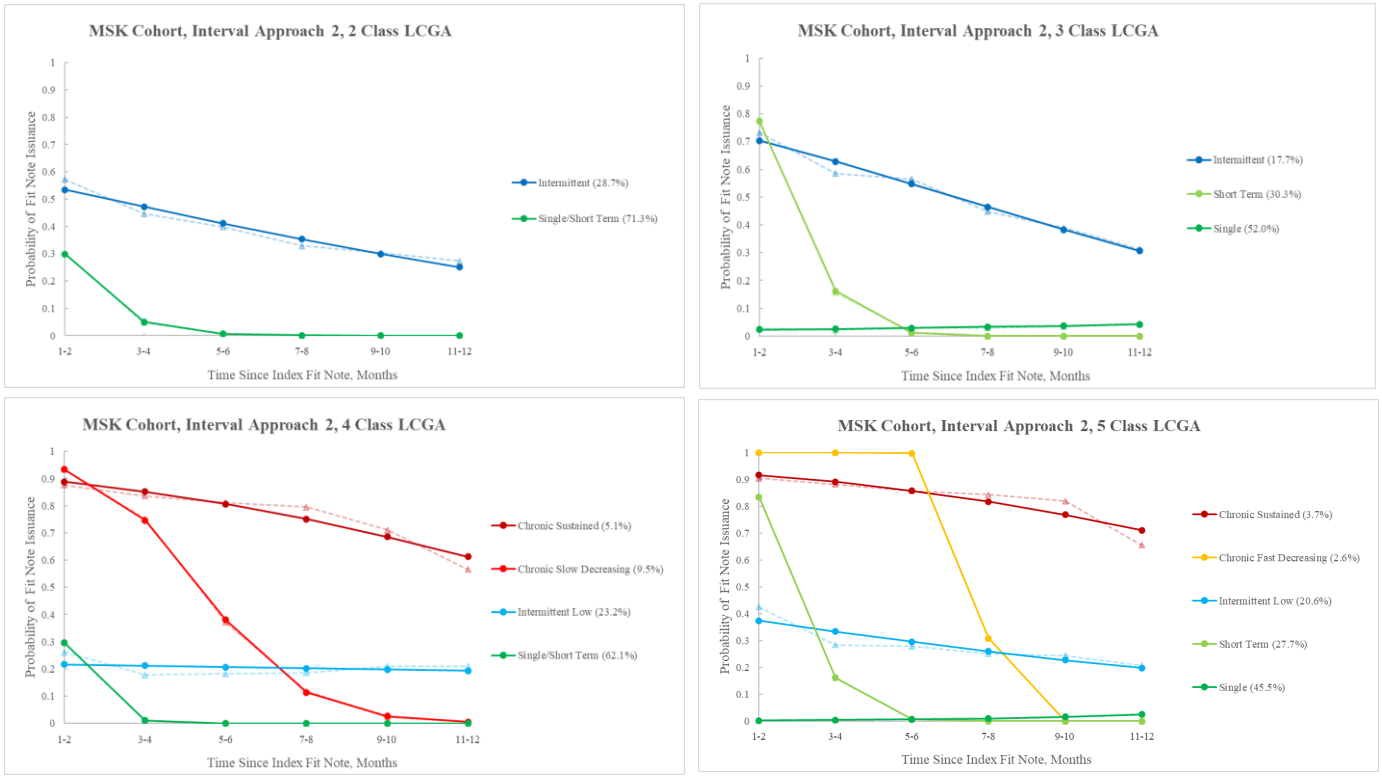
[a] LRT was assessed with both the Lo-Mendell-Rubin and bootstrapped methods

The two- and three-class LCGA models were highly similar to those observed for approach 1 in Section 7.3.1, in terms of the trajectory shapes (Figure 26), as well as the corresponding trajectory prevalences.

However, the four- and five-class models presented new trajectory classes.

In the four-class model, whilst the ‘Single/Short Term’ class remained as seen previously in the two-class LCGA models in approaches 1 and 2, the other three trajectories were novel. Firstly, one trajectory emerged whereby probability of fit note issuance remained constant and low at around 0.2, named ‘Intermittent Low’ (23.2% prevalence).

Figure 26. Linear LCGA Trajectory Model Plots for MSK Cohort, Interval Approach 2 (Two Monthly Intervals, Year One Data Only)



Abbreviations: LCGA = Latent Class Growth Analysis; MSK = Musculoskeletal.
For each trajectory, the solid lines represent the probability of fit note issuance in the given time interval (model estimated data), whilst the dotted lines represent the proportion of individuals issued a fit note (observed data). Trajectory class prevalence is shown, derived from a count based on posterior probability.

Secondly, two less favourable trajectories emerged. Both started with a high estimated probability of around 0.9 of fit note issuance in the first two months of follow-up. Then, in one trajectory class, named ‘Chronic Slow Decreasing’, there was a steady decrease in fit note issuance probability over time, reaching close to 0 at the end of one year follow-up (9.5% prevalence). Whilst in the least favourable trajectory class, ‘Chronic Sustained’ (5.1% prevalence), this high probability of fit note issuance in each two-month period was largely sustained for the first six months of follow-up, before decreasing gradually over the ensuing six months to 0.6 by month 11-12.

The five-class LCGA model was mostly similar to the four-class model, except that the ‘Single/Short Term’ trajectory from the four-class model appeared to have separated out into ‘Single’ and ‘Short Term’ trajectories (analogous to the change observed between the two- and three-class LCGA models described in the previous Section for approach 1).

Additionally, a new trajectory was identified in the five-class LCGA: ‘Chronic Fast Decreasing’ (2.6% prevalence), whereby the two-month probability of fit note issuance was high and close to 1 for the first six months of follow-up, before a rapid decrease occurred, and a probability close to 0 was reached and then sustained from month ten onwards.

Graphical assessment of model fit showed that the intermittent (both ‘Intermittent’ and ‘Intermittent Low’) and ‘Chronic Sustained’ trajectory fit note issuance estimates occasionally deviated from the observed proportions. The other trajectory classes had near perfect visual model fit.

7.3.3 Approach 3: Six Month Intervals (Years One to Three)

In this Section, models fitted based on the full three-year follow-up data are described, using six-monthly recurring intervals (interval approach 3).

Despite a longer-term recurring time interval being used here, the LGCM maintained a similar shape (Appendix Figure J.1) to the LGCMs based on approaches 1 and 2 (in Sections 7.3.1 and 7.3.2, respectively). The model fit indices suggested slight underperformance though (Appendix Table J.1). The CFI and TFI were now both below our guideline threshold of 0.95, with values of 0.94 and 0.92, respectively.

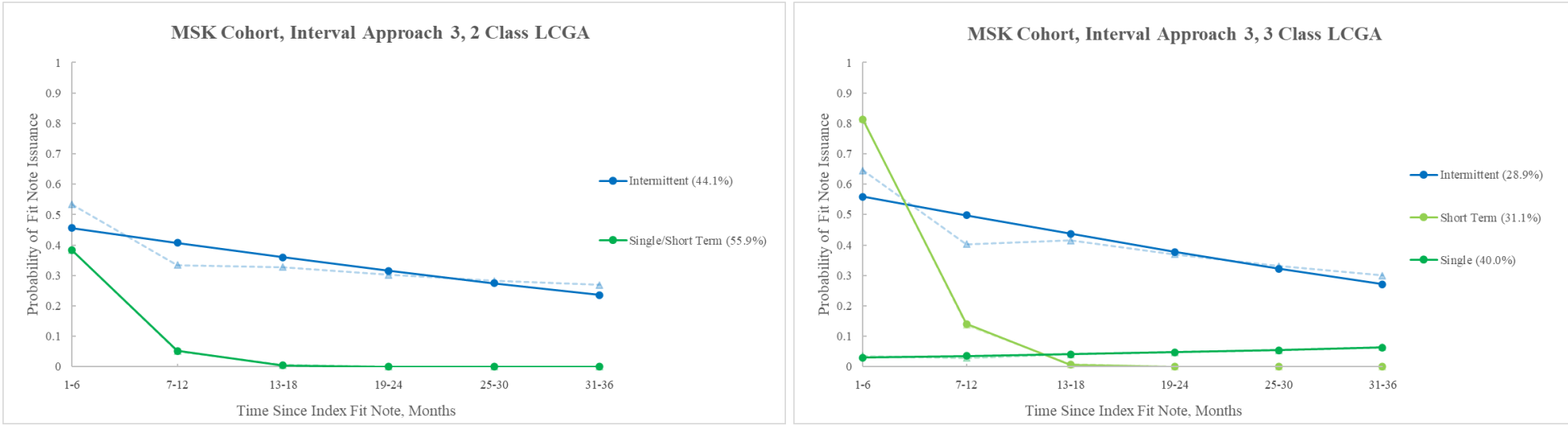
Similar to the three-monthly recurring interval approach 1 in Section 7.3.1, only LCGA models up to three classes converged when using six-monthly intervals (Table 25). The derived trajectories were also analogous to the corresponding two- and three-class models seen previously for approaches 1 and 2 (Figure 27).

Table 25. Model Fit and Class Meaningfulness Statistics of Linear LCGA Models for MSK Cohort, Interval Approach 3 (Six Monthly Intervals, Years One to Three Data)

	LCGA Model (n=42,222)	
	2 Class	**3 Class**
Log-likelihood	-102421	-101522
AIC	204852	203060
BIC	204895	203130
LRT[a]	p <0.0001	p <0.0001
Average Posterior Probability (Range)	0.86-0.92	0.80-0.88
Entropy	0.57	0.57
Trajectory Class Prevalences	55.9%, 44.1%	40.0%, 31.1%, 28.9%

Abbreviations: n = number of individuals; LCGA = Latent Class Growth Analysis; MSK = Musculoskeletal; AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; LRT = Likelihood Ratio Test.
[a] LRT was assessed with both the Lo-Mendell-Rubin and bootstrapped methods

Figure 27. Linear LCGA Trajectory Model Plots for MSK Cohort, Interval Approach 3 (Six Monthly Intervals, Years One to Three Data)



Abbreviations: LCGA = Latent Class Growth Analysis; MSK = Musculoskeletal.
For each trajectory, the solid lines represent the probability of fit note issuance in the given time interval (model estimated data), whilst the dotted lines represent the proportion of individuals issued a fit note (observed data). Trajectory class prevalence is shown, derived from a count based on posterior probability.

A notable difference, however, was that the graphical fit for the 'Intermittent' trajectory class, for both the two- and three-class models, now showed considerably greater divergence of estimated probabilities from the observed proportions. This divergence occurred predominantly during the first two-time intervals, i.e., the first year of follow-up. Thereafter, during years two and three of follow-up, the graphical model fit was near perfect.

The entropy was also lower than observed thus far with the previous two interval approaches used, with a value of 0.57 for both the two- and three-class models, respectively. These values were below our guideline threshold of 0.7, suggesting possible issues with class separability with approach 3.

7.3.4 Approach 4: Three Month Intervals (Year One); Six Month (Years Two to Three)

In this Section, the full three-year follow-up data was used once again, but now with shorter (three-monthly) recurring intervals used for year one, and the longer (six-monthly) recurring intervals restricted to years two and three.

The LGCM (Appendix Table J.1) had a similar RMSEA as from interval approach 3 (six-monthly recurring intervals), again meeting our model fit threshold. However, the CFI and TLI were also similar to interval approach 3, and again suggested underperformance. Furthermore, the SRMR deteriorated compared to interval approach 3 (at 0.07), which now didn't meet our threshold.

The shape of the LGCM (Appendix Figure J.1) was similar to the previously described LGCMs, except that there was now a small increase in probability of fit note issuance

from months thirteen to eighteen (compared to the probability between months nine and twelve), after which a plateau was reached.

However, one key difference with this interval approach compared to approaches 1 to 3, was that two-class LCGA did not run this time. Hence, in this instance, as described in Section 7.2.6, a more complex mixture model was explored: a piecewise mixture model.

Firstly, a piecewise LGCM was fit to the data. This improved all the model fit indices from the non-piecewise LGCM (Appendix Table J.1), such that all of these indices were now within our threshold of good performance. Visually, the differences between the non-piecewise and piecewise LGCMs were less pronounced (Appendix Figure J.1).

Next, piecewise LCGA models were fitted. Only the two-class model converged (Table 26 and Figure 28).

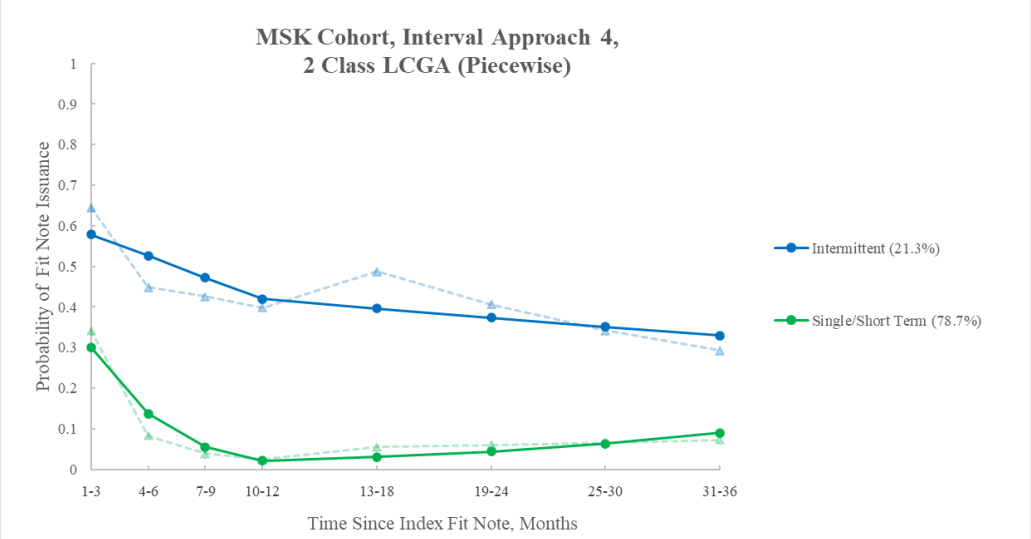
Table 26. Model Fit and Class Meaningfulness Statistics of Linear Piecewise Two-Class LCGA Model for MSK Cohort, Interval Approach 4 (Three Monthly Intervals in Year One, Six Monthly in Years Two to Three)

	LCGA Model (n=42,905)
	2 Class
Log-likelihood	-127647
AIC	255307
BIC	255368
LRT[a]	p <0.0001
Average Posterior Probability (Range)	0.86-0.93
Entropy	0.70
Trajectory Class Prevalences	78.7%, 21.3%

Abbreviations: n = number of individuals; LCGA = Latent Class Growth Analysis; MSK = Musculoskeletal; AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; LRT = Likelihood Ratio Test

[a] LRT was assessed with both the Lo-Mendell-Rubin and bootstrapped methods

Figure 28. Linear Piecewise Two-Class LCGA Trajectory Model Plot for MSK Cohort, Interval Approach 4 (Three Monthly Intervals in Year One, Six Monthly in Years Two to Three)



Abbreviations: LCGA = Latent Class Growth Analysis; MSK = Musculoskeletal.
For each trajectory, the solid lines represent the probability of fit note issuance in the given time interval (model estimated data), whilst the dotted lines represent the proportion of individuals issued a fit note (observed data). Trajectory class prevalence is shown, derived from a count based on posterior probability.

The entropy of this two-class LCGA model (0.7), was improved from approach 3, and now met our threshold criteria. Also, the two derived trajectories of this model were again similar to the two-class models from approaches 1-3. However, the graphical plot showed that the observed compared to estimated fit note issuance probabilities were misaligned in the ‘Intermittent’ trajectory class, and now this poor fit was observed during year two of follow-up data too (not only in year one of follow-up, as observed for this trajectory class when only six-monthly recurring intervals were used in approach 3).

7.3.5 Approach 5: Two Month Intervals (Year One); Six Month (Years Two to Three)

The final interval approach considered for trajectory derivation for individuals with an index fit note due to a MSK condition, was based on using yet shorter-term (two-monthly) recurring intervals in year one, and six-monthly in years two and three.

Model fit indices of the LGCM (Appendix Table J.1) were similar to the non-piecewise LGCM for approach 4 (in the previous Section), and thus indicative of possible poor fit again. Only the RMSEA met our threshold guidelines. Nonetheless, the graphical LGCM plot did not show any major concerns of model fit (Appendix Figure J.1).

Using two-monthly intervals for year one of follow-up data, instead of three-monthly (as in approach 4), led to the convergence of LCGA models with more classes – LCGA models with up to six classes are summarised in this Section (Table 27 and Figure 29). A seven-class LCGA could potentially have been explored too, but it was decided to stop at the six-class LCGA as some classes with low prevalence and weaker graphical model fit were identified.

The two trajectory classes identified from the two-class LCGA were largely similar to those from approach 4 (Section 7.3.4). Although, the 'Intermittent' class from approach 4 was slightly altered in approach 5 and generally had lower and sustained estimated probabilities of fit note issuance, of around 0.3 (and hence was named 'Intermittent Low').

Then, increasing the classes from two to three in the LCGA model, led to the same trajectories from the two-class model, with the addition of a 'Chronic Slow Decreasing' trajectory (11.5% prevalence) trajectory.

Table 27. Model Fit and Class Meaningfulness Statistics of Linear LCGA Models for MSK Cohort, Interval Approach 5 (Two Monthly Intervals in Year One, Six Monthly in Years Two to Three)

	LCGA Model (n=43,130)				
	2 Class	**3 Class**	**4 Class**	**5 Class**	**6 Class**
Log-likelihood	-152178	-147439	-145244	-144791	-144568
AIC	304366	294894	290511	289611	289170
BIC	304410	294963	290606	289732	289317
LRT[a]	p <0.0001	p <0.0001	p <0.0001	p <0.0001	p <0.0001
Average Posterior Probability (Range)	0.89-0.98	0.83-0.90	0.82-0.93	0.81-0.91	0.73-0.84
Entropy	0.71	0.66	0.73	0.75	0.71
Trajectory Class Prevalences	52.8%, 47.2%	49.7%, 38.8%, 11.5%	44.5%, 40.9%, 7.8%, 6.9%	43.9%, 41.3%, 7.3%, 5.1%, 2.5%	36.7%, 33.4%, 19.6%, 4.8%, 3.8%, 1.8%

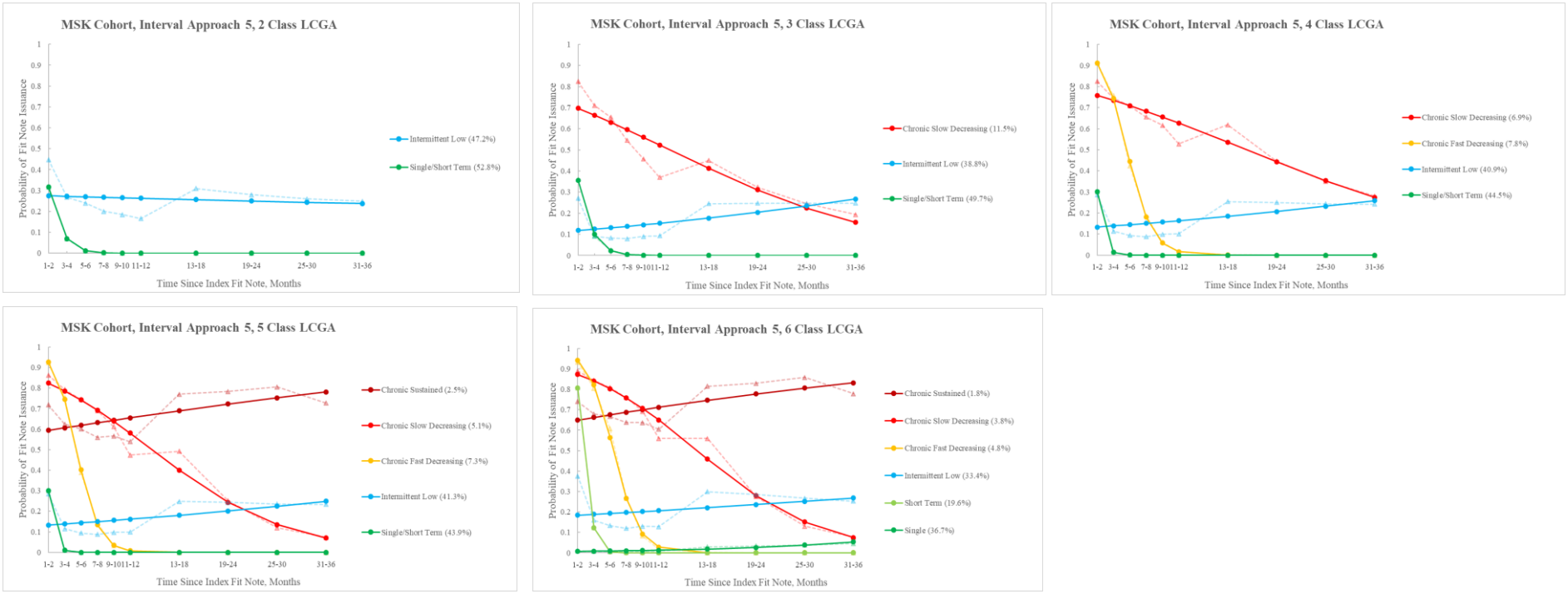
Abbreviations: n = number of individuals; LCGA = Latent Class Growth Analysis; MSK = Musculoskeletal; AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; LRT = Likelihood Ratio Test.
[a] LRT was assessed with both the Lo-Mendell-Rubin and bootstrapped methods

Increasing yet further to a four-class LCGA, the trajectories from the three-class LCGA in this Section were also retained, with the addition of a 'Chronic Fast Decreasing' class now (7.8% prevalence). Then, raising the classes to five further added a 'Chronic Sustained' trajectory (albeit with low prevalence of 2.5%).

Finally, in the six-class LCGA, four of the trajectories were retained as per the five-class model, except that the 'Single/Short Term' class was now split out into two separate classes (this effect was as observed in previous Sections).

The graphical assessment of model fit was weakest for the 'Intermittent Low', 'Chronic Slow Decreasing' and 'Chronic Sustained' classes, and near perfect for the other classes.

Figure 29. Linear LCGA Trajectory Model Plots for MSK Cohort, Interval Approach 5 (Two Monthly Intervals in Year One, Six Monthly in Years Two to Three)



Abbreviations: LCGA = Latent Class Growth Analysis; MSK = Musculoskeletal.
For each trajectory, the solid lines represent the probability of fit note issuance in the given time interval (model estimated data), whilst the dotted lines represent the proportion of individuals issued a fit note (observed data). Trajectory class prevalence is shown, derived from a count based on posterior probability.

Entropy values from all models in this Section were largely above our threshold guidelines, with the five-class LCGA having the highest value (0.75). The AIC and BIC model fit indices suggested that each LCGA outperformed the corresponding LCGA with one fewer class, albeit this improvement was smaller between the four to six class LCGA models.

In the next Section, the LGCM and LCGA results for the cohort with an index fit note due to a MH condition are presented, using the same structure as in this Section.

Then, in Section 7.5, an optimal LCGA model is chosen, separately for the index MSK and MH condition fit note cohorts, through taking into account all of the results from Sections 7.3 and 7.4. Once the optimal models are chosen for each cohort, GMM is explored for these final models too (also in Section 7.5).

7.4 Results: Trajectories of Work Absence Due to a MH Condition

7.4.1 Approach 1: Three Month Intervals (Year One Only)

In line with the results from approach 1 for individuals with an index MSK condition fit note, the LGCM under approach 1 for the index MH condition fit note cohort also exhibited good model fit (Appendix Table J.2 and Figure J.2). All of our threshold criteria were met. The LGCM plots for index MH condition individuals also had a similar shape to the LGCM plots for MSK approach 1, except that the starting probability of fit note issuance in the first three-month interval was slightly higher for index MH condition individuals, ranging between around 0.5 to 0.6.

As in MSK approach 1, the LCGA models for MH approach 1 also only converged up to three-classes (Table 28 and Figure 30). The derived LCGA trajectories were similar in nature to those from approach 1 from analysis of MSK fit notes, except that the initial

estimated probabilities of fit note issuance tended to be higher for index MH condition individuals (in line with the same observation mentioned above for the LGCM). For example, in the three-class LCGA 'Short Term' trajectory under approach 1, the probability of fit note issuance in the first three months of follow-up was 0.83 for index MSK condition individuals, and 0.95 for index MH condition individuals.

In terms of overall performance of the LCGA models under interval approach 1 (comparing both the MSK and MH condition fit note analysis), the three-class LCGA model from the MH condition fit note cohort was the only one that had an entropy above our guideline threshold value (0.73).

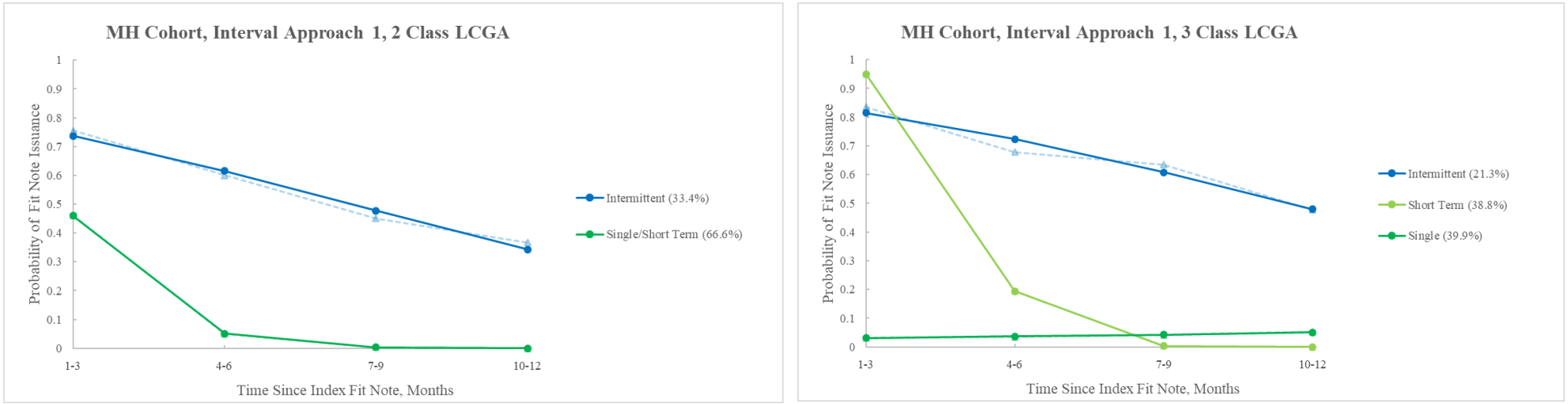
Table 28. Model Fit and Class Meaningfulness Statistics of Linear LCGA Models for MH Cohort, Interval Approach 1 (Three Monthly Intervals, Year One Data Only)

	LCGA Model (n=61,900)	
	2 Class	**3 Class**
Log-likelihood	-113231	-111345
AIC	226472	222707
BIC	226517	222779
LRT[a]	p <0.0001	p <0.0001
Average Posterior Probability (Range)	0.89-0.92	0.88-0.96
Entropy	0.64	0.73
Trajectory Class Prevalences	66.6%, 33.4%	39.9%, 38.8%, 21.3%

Abbreviations: n = number of individuals; LCGA = Latent Class Growth Analysis; MH = Mental Health; AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; LRT = Likelihood Ratio Test.

[a] LRT was assessed with both the Lo-Mendell-Rubin and bootstrapped method

Figure 30. Linear LCGA Trajectory Model Plots for MH Cohort, Interval Approach 1 (Three Monthly Intervals, Year One Data Only)



Abbreviations: LCGA = Latent Class Growth Analysis; MH = Mental Health.
For each trajectory, the solid lines represent the probability of fit note issuance in the given time interval (model estimated data), whilst the dotted lines represent the proportion of individuals issued a fit note (observed data). Trajectory class prevalence is shown, derived from a count based on posterior probability.

7.4.2 Approach 2: Two Month Intervals (Year One Only)

When two-monthly recurring intervals were used, the LGCM exhibited slightly improved fit compared to approach 1 (Appendix Table J.2 and Figure J.2), in line with observations from the LGCMs from approaches 1 and 2 under the MSK condition fit note analysis.

In particular, the RMSEA improved from 0.07 to 0.05, when using interval approach 2 instead of approach 1, respectively, and the TLI from 0.95 (which was exactly at our threshold cut-off value) to 0.97 (above the threshold value).

LCGA models with up to six-classes all converged (Table 29 and Figure 31). A seven-class LCGA could have been explored here, but six-classes were deemed sufficient due to the identification of some trajectories with low prevalences (2.7% and 2.9% were the two lowest prevalences). Furthermore, the AIC and BIC values, whilst demonstrating that LCGA models with progressively higher classes were improvements on the same model with one fewer class, showed less difference between the four to six class models. Hence if a seven-class model was explored, negligible improvement in model performance might have been expected.

The two- to five-class LCGA models were as described for index MSK condition individuals under approach 2 in Section 7.3.2.

But, in the six-class LCGA, two novel trajectory classes were identified, the first was named ‘Chronic Long, Fast Decreasing’ (2.7% prevalence). This class was similar to the ‘Chronic Fast Decreasing’ trajectory (10.0% prevalence) that was also present in this six-class LCGA, yet differentiated by individuals having a longer sustained probability close to 1 of fit note issuance (up to month eight in this ‘Chronic Long, Fast

Decreasing' class, compared to up to month four in the 'Chronic Fast Decreasing' class). After this point, in both trajectory classes there was the same rapid decrease in probability of fit note issuance.

The second novel trajectory class from the six-class LCGA was named 'Single with Later Relapse' (2.9% prevalence). Individuals in this trajectory maintained a two-month probability close to 0 of fit note issuance over the first six months of follow-up. However, there was a steady increase in fit note issuance probability from month six to the end of year one of follow-up (reaching 0.5).

Entropy and average posterior probabilities for all the LCGA models in this Section were generally above our guideline threshold.

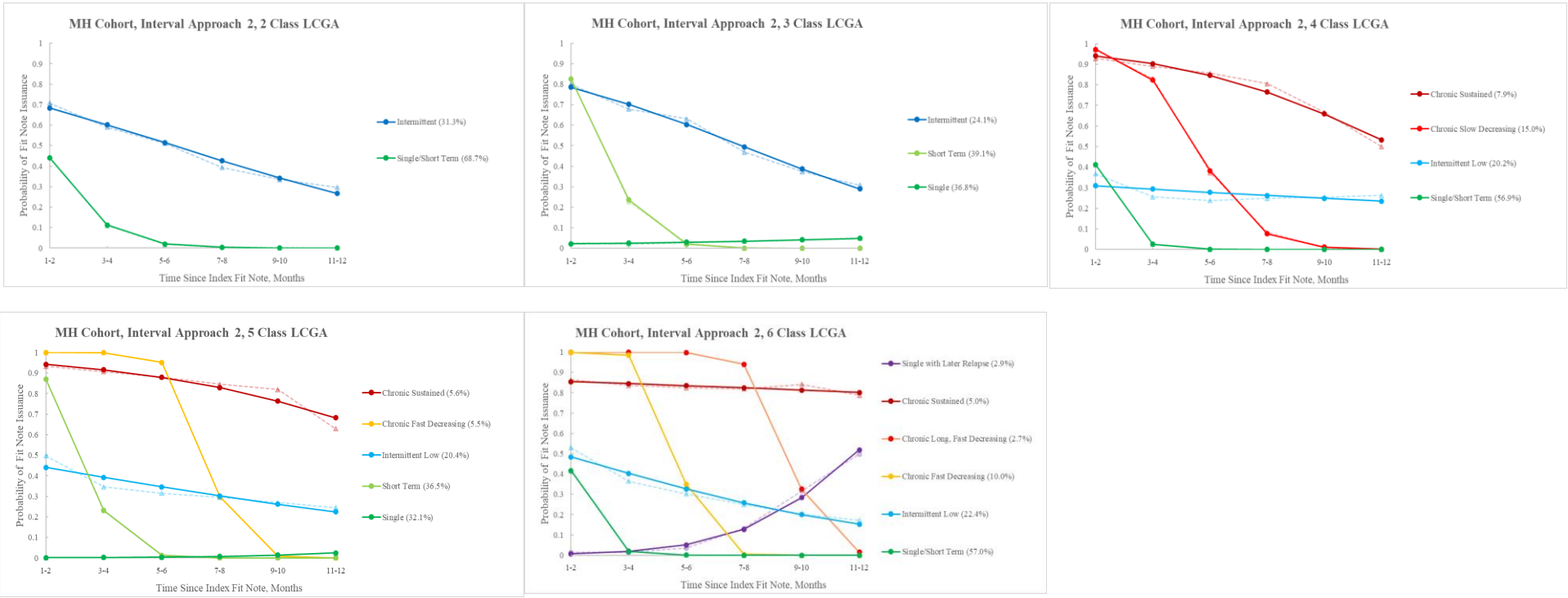
Table 29. Model Fit and Class Meaningfulness Statistics of Linear LCGA Models for MH Cohort, Interval Approach 2 (Two Monthly Intervals, Year One Data Only)

	LCGA Model (n=62,355)				
	2 Class	**3 Class**	**4 Class**	**5 Class**	**6 Class**
Log-likelihood	-149801	-146570	-144110	-143694	-143520
AIC	299612	293155	288242	287415	287073
BIC	299657	293228	288341	287542	287227
LRT[a]	p <0.0001	p <0.0001	p <0.0001	p <0.0001	p <0.0001
Average Posterior Probability (Range)	0.92-0.96	0.85-0.93	0.83-0.89	0.81-0.91	0.68-0.90
Entropy	0.73	0.69	0.71	0.73	0.77
Trajectory Class Prevalences	68.7%, 31.3%	39.1%, 36.8%, 24.1%	56.9%, 20.2%, 15.0%, 7.9%	36.5%, 32.1%, 20.4%, 5.6%, 5.5%	57.0%, 22.4%, 10.0%, 5.0%, 2.9%, 2.7%

Abbreviations: n = number of individuals; LCGA = Latent Class Growth Analysis; MH = Mental Health; AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; LRT = Likelihood Ratio Test.

[a] LRT was assessed with both the Lo-Mendell-Rubin and bootstrapped method

Figure 31. Linear LCGA Trajectory Model Plots for MH Cohort, Interval Approach 2 (Two Monthly Intervals, Year One Data Only)



Abbreviations: LCGA = Latent Class Growth Analysis; MH = Mental Health.
For each trajectory, the solid lines represent the probability of fit note issuance in the given time interval (model estimated data), whilst the dotted lines represent the proportion of individuals issued a fit note (observed data). Trajectory class prevalence is shown, derived from a count based on posterior probability

7.4.3 Approach 3: Six Month Intervals (Years One to Three)

The LGCM as applied to six-monthly recurring intervals over the full three-year follow-up data yielded a poor fitting model (Appendix Table J.2 and Figure J.2). Only the RMSEA was within our guideline threshold. The CFI was 0.91, and the TLI was especially low at 0.88. The SRMR was high at 0.07. These model fit index values were worse than for the LGCM under approach 3 in the MSK condition fit note analysis (Section 7.3.3).

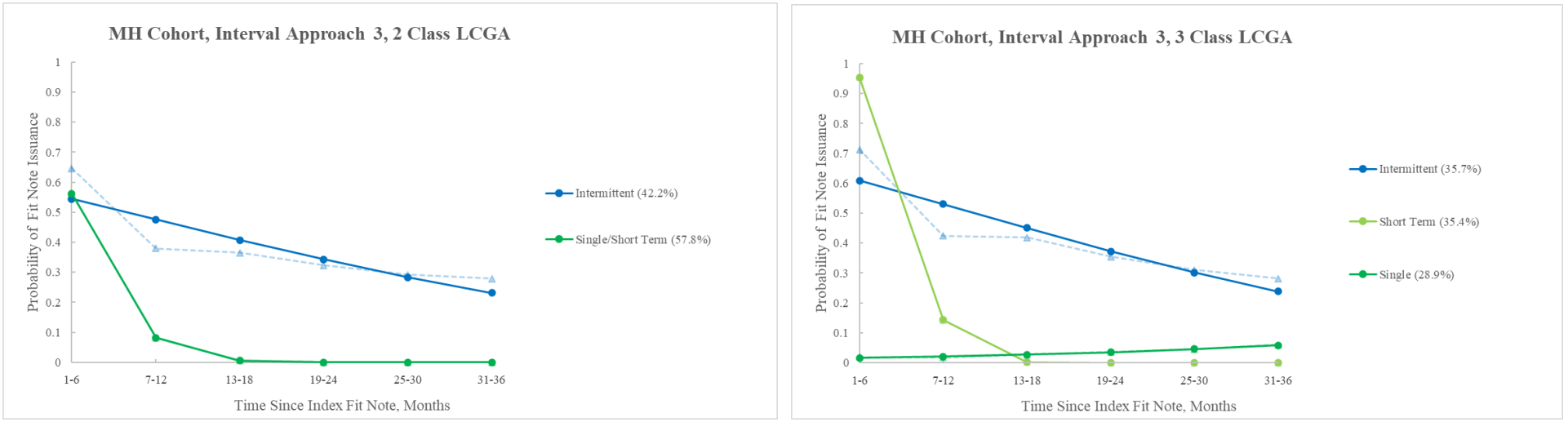
However, similar to MSK approach 3 (Section 7.3.3), LCGA models up to three-classes converged, and the derived trajectories were also very similar to those for individuals with an index MSK condition (Table 30 and Figure 32).

Table 30. Model Fit and Class Meaningfulness Statistics of Linear LCGA Models for MH Cohort, Interval Approach 3 (Six Monthly Intervals, Years One to Three Data)

	LCGA Model (n=60,536)	
	2 Class	**3 Class**
Log-likelihood	-147696	-146465
AIC	295402	292947
BIC	295447	293019
LRT[a]	p <0.0001	p <0.0001
Average Posterior Probability (Range)	0.84-0.97	0.85-0.93
Entropy	0.59	0.68
Trajectory Class Prevalences	57.8%, 42.2%	35.7%, 35.4%, 28.9%

Abbreviations: n = number of individuals; LCGA = Latent Class Growth Analysis; MH = Mental Health; AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; LRT = Likelihood Ratio Test.
[a] LRT was assessed with both the Lo-Mendell-Rubin and bootstrapped methods

Figure 32. Linear LCGA Trajectory Model Plots for MH Cohort, Interval Approach 3 (Six Monthly Intervals, Years One to Three Data)



Abbreviations: LCGA = Latent Class Growth Analysis; MH = Mental Health.
For each trajectory, the solid lines represent the probability of fit note issuance in the given time interval (model estimated data), whilst the dotted lines represent the proportion of individuals issued a fit note (observed data). Trajectory class prevalence is shown, derived from a count based on posterior probability.

Yet, the entropy of the LCGA models derived in this Section were higher than the corresponding values for MSK approach 3 (in Section 7.3.3), at 0.59 and 0.68, for the two- and three-class LCGA models, respectively (compared to 0.57 in both models for MSK approach 3). Albeit these entropy values were still below our guideline threshold.

7.4.4 Approach 4: Three Month Intervals (Year One); Six Month (Years Two to Three)

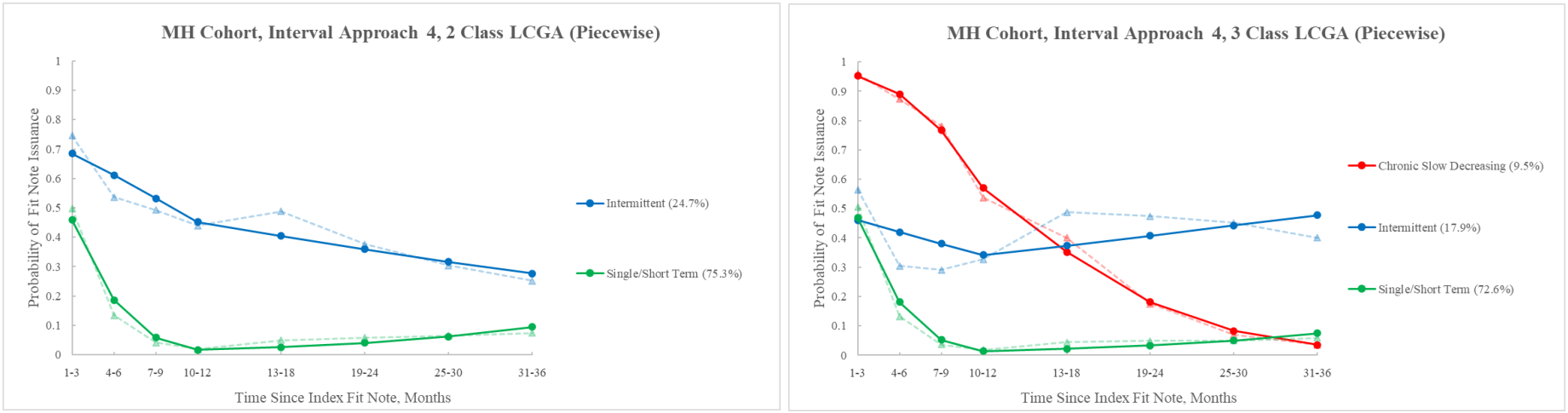
When three-monthly recurring intervals were used for year one of follow-up data, and six-monthly intervals thereafter, the LGCM model fit remained poor, as in approach 3 (Appendix Table J.2 and Figure J.2). The model fit indices were similar to those for approach 3 in the previous Section, with the CFI, TLI and SRMR all outside of our guideline threshold values. In particular, the SRMR reached its joint highest value in this study of 0.09.

Similar to the MSK condition fit note analysis under approach 4, the conventional (non-piecewise) two-class LCGA also did not converge, hence piecewise models were fitted (including a piecewise LGCM for completeness). The piecewise LGCM improved all of the model fit indices of the non-piecewise LGCM, with all except the TLI measure now meeting our guideline thresholds.

Then, piecewise LCGA models up to three classes converged (Table 31 and Figure 33).

The derived trajectories of the two-class piecewise LCGA, ‘Intermittent’ and ‘Single/Short Term’ were analogous to the two-class LCGA trajectories under approaches 1-3 from MH condition fit note analysis. Except, that the visual fit of estimated fit note issuance probabilities of the ‘Intermittent’ class was misaligned compared to the observed fit note proportions (as with approach 3 only).

Figure 33. Linear Piecewise LCGA Trajectory Model Plots for MH Cohort, Interval Approach 4 (Three Monthly Intervals in Year One, Six Monthly in Years Two to Three)



Abbreviations: LCGA = Latent Class Growth Analysis; MH = Mental Health.
For each trajectory, the solid lines represent the probability of fit note issuance in the given time interval (model estimated data), whilst the dotted lines represent the proportion of individuals issued a fit note (observed data). Trajectory class prevalence is shown, derived from a count based on posterior probability

Table 31. Model Fit and Class Meaningfulness Statistics of Linear Piecewise LCGA Models for MH Cohort, Interval Approach 4 (Three Monthly Intervals in Year One, Six Monthly in Years Two to Three)

	LCGA Model (n=61,900)	
	2 Class	**3 Class**
Log-likelihood	-193049	-190826
AIC	386113	381674
BIC	386176	381773
LRT[a]	p <0.0001	p <0.0001
Average Posterior Probability (Range)	0.86-0.93	0.78-0.92
Entropy	0.69	0.72
Trajectory Class Prevalences	75.3%, 24.7%	72.6%, 17.9%, 9.5%

Abbreviations: n = number of individuals; LCGA = Latent Class Growth Analysis; MH = Mental Health; AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; LRT = Likelihood Ratio Test.
[a] LRT was assessed with both the Lo-Mendell-Rubin and bootstrapped methods

Finally, in the three-class LCGA, a third trajectory was added to the two-class LCGA, 'Chronic Slow Decreasing'. The entropy of the two- and three-class LCGAs was at the level of our guideline threshold.

7.4.5 Approach 5: Two Month Intervals (Year One); Six Month (Years Two to Three)

The final set of results presented in this study, are for two-monthly recurring intervals in year one of follow-up, and six-monthly recurring intervals for years two and three. The LGCM model fit was slightly improved compared to that of the non-piecewise LGCM under approach 4 in the previous Section (Appendix Table J.2 and Figure J.2), although did not meet our threshold guidelines (except for the RMSEA value).

LCGA models up to six-classes converged (Table 32 and Figure 34), as for approach 5 in the MSK condition fit note analysis (Section 7.3.5), and the class shapes and

trajectory prevalences were also similar. The same reasoning to stop at six-classes and not explore a seven-class LCGA was applied here too.

Entropy was largely above our guideline threshold for all LCGA models, and in-line with MSK approach 5.

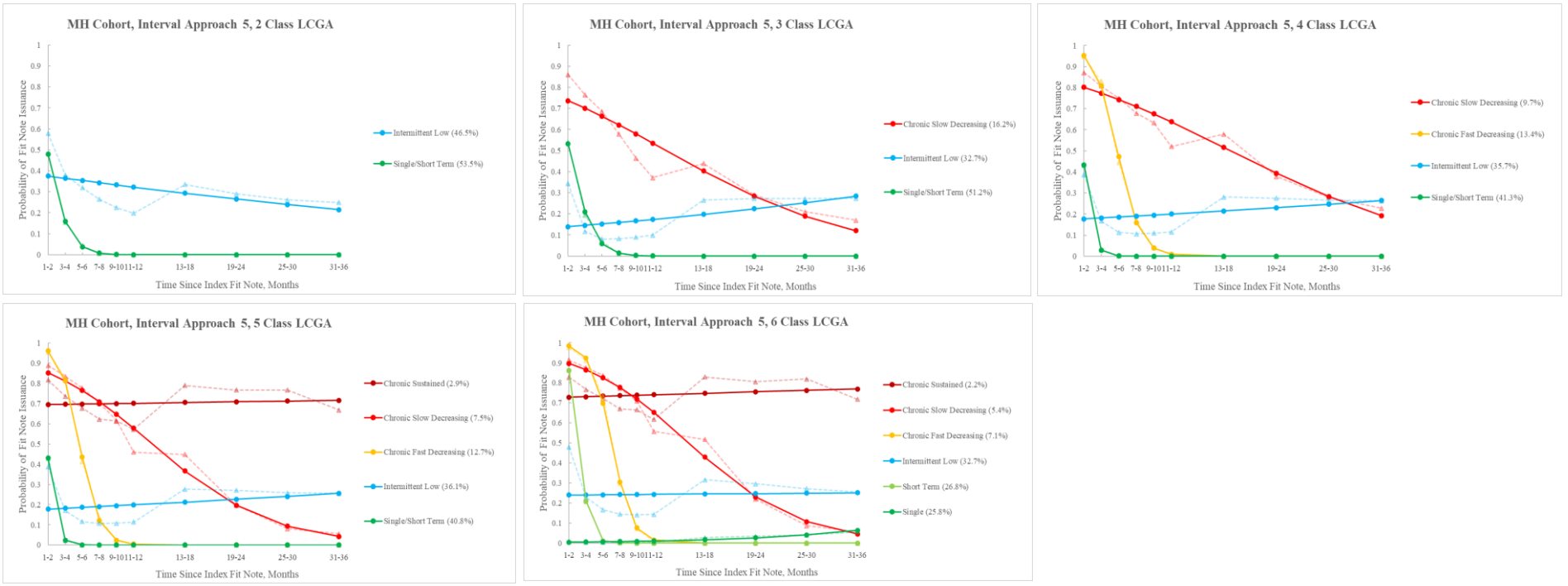
Table 32. Model Fit and Class Meaningfulness Statistics of Linear LCGA Models for MH Cohort, Interval Approach 5 (Two Monthly Intervals in Year One, Six Monthly in Years Two to Three)

	LCGA Model (n=62,355)				
	2 Class	**3 Class**	**4 Class**	**5 Class**	**6 Class**
Log-likelihood	-240427	-232552	-227594	-226698	-226149
AIC	480865	465120	455210	453424	452332
BIC	480910	465192	455309	453551	452486
LRT[a]	p <0.0001	p <0.0001	p <0.0001	p <0.0001	p <0.0001
Average Posterior Probability (Range)	0.91-0.95	0.84-0.88	0.84-0.90	0.80-0.89	0.77-0.87
Entropy	0.73	0.67	0.73	0.75	0.73
Trajectory Class Prevalences	53.5%, 46.5%	51.2%, 32.7%, 16.2%	41.3%, 35.7%, 13.4%, 9.7%	40.8%, 36.1%, 12.7%, 7.5%, 2.9%	32.7% 26.8%, 25.8%, 7.1%, 5.4%, 2.2%

Abbreviations: n = number of individuals; LCGA = Latent Class Growth Analysis; MH = Mental Health; AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; LRT = Likelihood Ratio Test.

[a] LRT was assessed with both the Lo-Mendell-Rubin and bootstrapped methods

Figure 34. Linear LCGA Trajectory Model Plots for MH Cohort, Interval Approach 5 (Two Monthly Intervals in Year One, Six Monthly in Years Two to Three)



Abbreviations: LCGA = Latent Class Growth Analysis; MH = Mental Health.
For each trajectory, the solid lines represent the probability of fit note issuance in the given time interval (model estimated data), whilst the dotted lines represent the proportion of individuals issued a fit note (observed data). Trajectory class prevalence is shown, derived from a count based on posterior probability.

7.5 Results: Optimal Trajectory Model

In this Section, an optimal LCGA trajectory model is first chosen (Section 7.5.1), separately for the cohort with an index MSK condition fit note, and with an index MH condition fit note.

Then application of a GMM is tested for these two optimal LCGA models, to assess whether this model type should replace the optimal model (Section 7.5.2).

Following this, the impact of the two sensitivity analyses for different approaches to handling missing data is described in Section 7.5.3.

Finally, the chosen optimal model (either LCGA or GMM), is summarised, and individual variability within the trajectory classes is explored (Section 7.5.4).

7.5.1 Choosing an Optimal Trajectory LCGA Model

As will be explained in this Section, a five-class LCGA model using interval approach 2 (one year follow-up with two-monthly intervals) was chosen as the optimal LCGA model, for both the MSK and MH condition fit note cohorts.

This decision was reached taking into consideration the trajectory derivation results across all five interval approaches from Sections 7.3 and 7.4, and our model assessment guidelines from Table 9.

Furthermore, this decision was made through consensus between AL and his supervisory team, as well as via incorporating feedback from:

- A GP (AL presented his fit note trajectory derivation results to a GP on 25th May 2023)

- OHID, who also agreed with this choice after AL presented his decision-making process to them on 21st June 2023
- A PPIE group, who affirmed the plausibility and validity of these five types of absence trajectory based on their personal experiences of sickness absence (AL led a PPIE session 1st February 2024 and explained the absence behaviour identified in the five different subgroups of the optimal models)

In the first step of the decision-making process, the general performance of the LCGA models under each of the five interval approaches was first assessed (summarised in Table 33 and Table 34, for the MSK and MH condition fit note analyses, respectively).

For both the MSK and MH condition fit note trajectory derivation analysis, interval approaches 1, 3, and 4 (involving either three- and/or six-monthly recurring intervals) performed worst. These three interval approaches exhibited the lowest entropy values, either close to or below our guideline threshold of 0.7. In some cases, the entropy was considerably lower than the guideline threshold (for example, the highest entropy value under approach 3 was 0.57 for the index MSK condition fit note cohort), suggesting possible class separability issues. In contrast, the entropy values of the LCGA models fit under approaches 2 and 5 (which featured two-monthly intervals) were higher.

Furthermore, there was a lack of variability in the derived trajectories based on interval approaches 1, 3 and 4 - only a maximum of three classes were uncovered. Whilst under interval approaches 2 and 5, not only were the trajectories from approaches 1, 3 and 4 replicated, but additional trajectory classes were also uncovered.

Table 33. Summary of LCGA Trajectory Models Based on Individuals with an Index Fit Note Due to a MSK Condition

Interval Approach	Year One Follow-Up Time Interval	Years Two to Three Follow-Up Time Interval	Highest Number of Classes Used in LCGA Models	Highest Entropy Value
1	Three-Monthly	-	3	0.66
2	Two-Monthly	-	5	0.72
3	Six-Monthly	Six-Monthly	3	0.57
4	Three-Monthly	Six-Monthly	2[a]	0.70
5	Two-Monthly	Six-Monthly	6	0.75

Abbreviations: LCGA = Latent Class Growth Analysis; MSK = Musculoskeletal

[a] This is based on piecewise LCGA, as the non-piecewise two-class LCGA did not run successfully under this interval approach

Table 34. Summary of LCGA Trajectory Models Based on Individuals with an Index Fit Note Due to a MH Condition

Interval Approach	Year One Follow-Up Time Interval	Years Two to Three Follow-Up Time Interval	Highest Number of Classes Used in LCGA Models	Highest Entropy Value
1	Three-Monthly	-	3	0.73
2	Two-Monthly	-	6	0.77
3	Six-Monthly	Six-Monthly	3	0.68
4	Three-Monthly	Six-Monthly	3[a]	0.72
5	Two-Monthly	Six-Monthly	6	0.75

Abbreviations: LCGA = Latent Class Growth Analysis; MH = Mental Health

[a] This is based on piecewise LCGA, as the non-piecewise two-class LCGA did not run successfully under this interval approach

Therefore, due to inferior class separability performance and fewer derived trajectories, LCGA models relating to approaches 1, 3 and 4 were excluded from consideration for the optimal model.

Then, of the two remaining interval approaches (2 and 5), approach 5 was also ruled out. The rationale for this, first for the MSK condition index fit note cohort, was through comparing the LCGA plots from approach 5 (Figure 29) that are based on three-year follow-up, to those of approach 2 (Figure 26) based on one year follow-up. Whilst there were some similarities in the classes derived, a notable difference was that the trajectories from approach 5 were more prone to poor graphical fit compared to the observed data, whereas those of approach 2 generally showed good fit. The same trend was observed for the MH condition index fit note cohort (comparing the plots from approach 5 - Figure 34, to those of approach 2 - Figure 31).

Therefore, approach 2 (two-monthly recurring intervals for year one data only) was considered the optimal interval approach in this study, for both the MSK and MH index fit note cohorts.

Then, first considering the index MSK condition fit note cohort, the optimal model choice was narrowed down to either the four- or five-class LCGA within approach 2. Both of these models exhibited good performance, with:

- The lowest and similar AIC and BIC values
- Statistically significant *p* values from the LMR-LRT and BLRT tests

These model fit index values suggested good model fit, which was further affirmed from the trajectory class plots (Figure 26), albeit the ‘Intermittent Low’ and ‘Chronic

Sustained' classes (present in both the four- and five-class LCGA models) showed some instances of poor visual model fit.

Additionally, meaningfulness of classes in these four- and five-class LCGA models was demonstrated by:

- Entropy and average posterior probability values ≥0.7 for all classes, suggesting good class separability
- Minimum prevalence ≥1% for all classes, thereby reducing the risk of spurious results being discovered
- A variety of different and plausible trajectory class shapes, that made sense in a work absence context

Ultimately, for the MSK condition index fit note cohort, it was decided that the five-class LCGA under interval approach 2 was the most appropriate choice (Figure 35). This decision was made considering clinical relevance and plausibility of the classes. In particular, the five-class LCGA provided a clear distinction between the subgroups who had a 'Short Term' sickness absence episode, compared to a 'Single' index fit note, unlike the four-class LCGA. A GP also confirmed that these two 'Short Term' and 'Single' subgroups were common in her experience of issuing fit notes.

Additionally, the 'Chronic Fast Decreasing' class that was present in the five-class LCGA, and not in the four-class model, seemed reasonable in the context of UK sick pay. SSP in the UK, is available for eligible claimants for a period of up to twenty-eight weeks.[5] After this point, additional remuneration may be claimable following another evaluation. For example, an application can be made to claim benefits such as the ESA or Universal Credit.[5] Therefore, the sharp decrease in probability of fit note issuance

after six months of follow-up observed in this trajectory class may be influenced by the original SSP period ending.

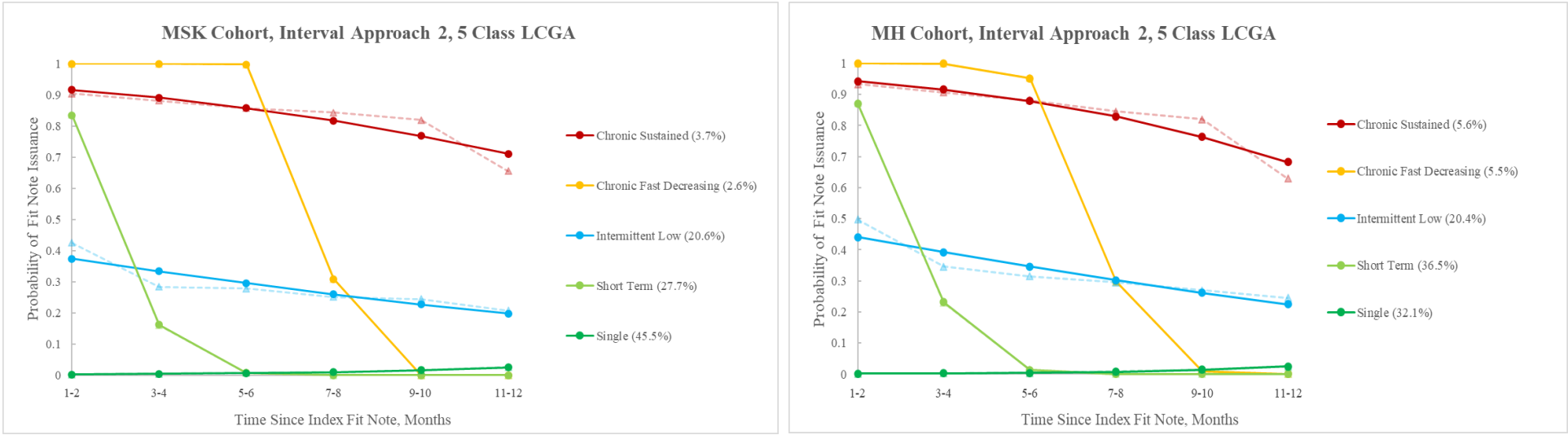
Finally, for the index MH condition fit note cohort, from the two-monthly recurring interval LCGAs under approach 2, the five-class LCGA was again chosen as the optimal model (Figure 35). In the decision-making process, the four-, five- and six-class LCGAs were all contenders for the optimal model, for the same reasons as given for models based on incidence absence due to a MSK condition.

In particular, the six-class model contained the previously mentioned novel ‘Single with Later Relapse’ class (2.9% prevalence). This ‘Single with Later Relapse’ class was also a subgroup of clinical interest highlighted by the aforementioned GP. Furthermore, this six-class LCGA differentiated between two types of ‘Chronic Fast Decreasing’ class, with a novel ‘Chronic Long, Fast Decreasing’ class (2.7% prevalence).

However, both of these two classes occurred with low prevalences, and were not observed in any other LCGA models in this study. Hence there was a possibility that the uncovering of these two novel classes could be spurious findings.

Additionally, in consideration of parsimony, as well as consistency with the MSK fit note final model, it was decided to reject the six-class model and choose the five-class LCGA as the optimal model for the MH fit note analysis (Figure 35).

Figure 35. Optimal Five-Class LCGA Models, Based on Interval Approach 2 (Two Monthly Intervals, Year One Data Only), for the Cohort with Index Fit Note Due to a MSK Condition (Left) and MH Condition (Right)



Abbreviations: LCGA = Latent Class Growth Analysis; MSK = Musculoskeletal; MH = Mental Health.
For each trajectory, the solid lines represent the probability of fit note issuance in the given time interval (model estimated data), whilst the dotted lines represent the proportion of individuals issued a fit note (observed data). Trajectory class prevalence is shown, derived from a count based on posterior probability.

7.5.2 GMM Considerations

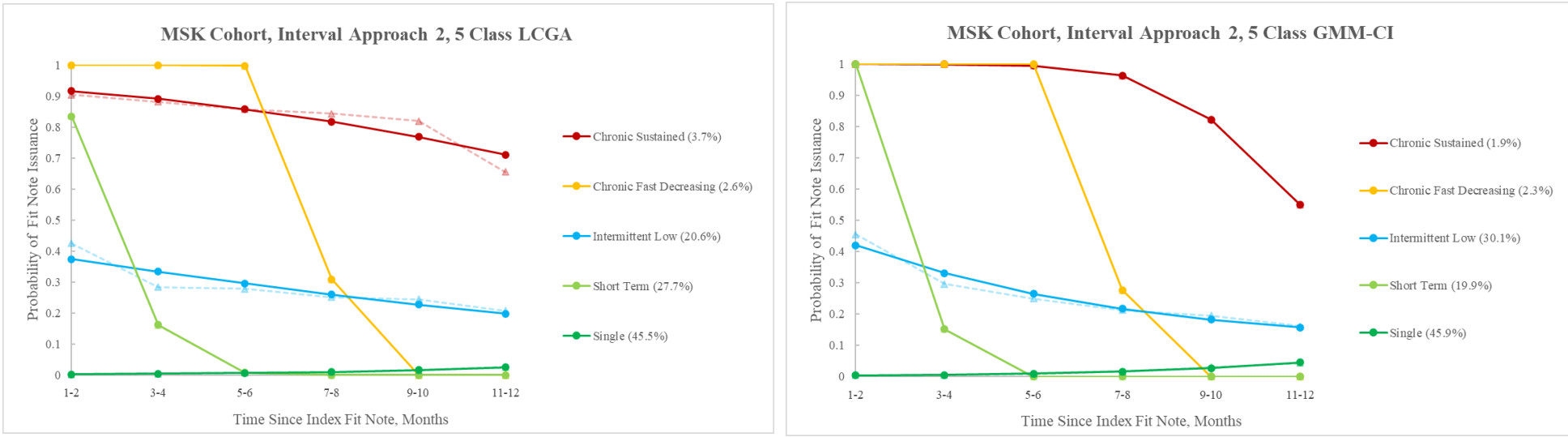
Next, to test if using a GMM approach would improve the optimal trajectory models, a five-class GMM-CI model based on interval approach 2 was first explored, initially for the incident MSK condition fit note cohort.

After a run time of 50 hours and 12 minutes (in comparison, the optimal five-class LCGA took 26 minutes to run), whilst there was no error message concerning model convergence, it was not clear whether a local maxima solution had indeed been achieved, rather than a global solution. Upon inspection of the output, 85 perturbed starting value runs did not converge or were rejected, and by examining the ordered list of final stage loglikelihood values, the top 3 values were similar, but not identical (as they were in all the LCGA models of this study): -87282.645, -87282.656, -87282.703.

Furthermore, the model output produced an error message stating that the standard errors of the model parameter estimates might not be trustworthy due to a non-positive definite first-order derivative matrix. It was suggested that this may be due to the starting values or an indication of model nonidentification.

Hence, due to these issues with the five-class GMM CI model under interval approach 2 potentially not converging to a global maximum loglikelihood solution, and possibly having unreliable solutions, the results from this model must be taken with caution. Nonetheless, as can be seen below in Figure 36, the derived trajectories from the optimal five-class LCGA model and the five-class GMM-CI for the index MSK condition fit note cohort under interval approach 2 showed similarity in shape and prevalence. In particular, the ‘Single’ and ‘Chronic Fast Decreasing’ trajectory shapes and prevalences were almost identical across the two approaches.

Figure 36. Five-Class Models for the Cohort with Index Fit Note Due to a MSK Condition, Based on Interval Approach 2 (Two Monthly Intervals, Years One to Three Data), Based on LCGA (Left) and GMM-CI (Right)



Abbreviations: LCGA = Latent Class Growth Analysis; GMM-CI = Growth Mixture Modelling, using a Class Invariant approach; MSK = Musculoskeletal.
For each trajectory, the solid lines represent the probability of fit note issuance in the given time interval (model estimated data), whilst the dotted lines represent the proportion of individuals issued a fit note (observed data). Trajectory class prevalence is shown, derived from a count based on posterior probability.

The four-class GMM CI under interval approach 2 was also attempted, although there was no error message concerning standard errors now, there were similar concerns regarding model convergence. As there were no explicit error messages for the four-class GMM-CI model, a four-class GMM-CV was attempted for completeness. With this four-class GMM-CV model under interval approach 2 there was a technical error message concerning model convergence, that either model estimation had reached a saddle point or that the observed and expected information matrices did not match.

Due to the four- and five-class GMM models under interval 2 presenting issues, no further GMM models were considered in this study.

7.5.3 Sensitivity Analyses

In this Section, the two alternative approaches to handling missing data are revisited:

- Sensitivity analyses 1: individuals were excluded if they met an exclusion reason at any point during the calendar year of the index fit note, without any censoring after index year. This was performed for all analyses (interval approaches 1 to 5, for both the index MSK and MH condition fit note cohorts)
- Sensitivity analyses 2: individuals were excluded if they met an exclusion reason at any point during follow-up. This was performed only for the optimal trajectory models. As the optimal models were based on interval approach 2, this sensitivity analysis thus excluded individuals who did not have one complete year of follow-up from index fit note.

Comparing trajectory derivation results from the main analyses to sensitivity analyses 1 did not result in any notable differences (in terms of the shapes and prevalences of the trajectories, as well as performance of the models) across any interval approach within

the index MSK or MH condition fit note cohorts. Equally, the main analyses results were similar to those of sensitivity analyses 2, when comparing the optimal five-class LCGA models under interval approach 2.

To illustrate the similarity of results across the main approach and the two sensitivity analyses, trajectory plots for each of these three analysis approaches are shown for the optimal five-class LCGA model under interval approach 2, for the index MSK and MH condition fit note cohorts in Figure 37 and Figure 38, respectively, and the statistical performance of these models is summarised in Table 35 and Table 36, respectively.

Table 35. Comparison of Optimal LCGA Model Fit and Class Meaningfulness Statistics for MSK Cohort, Based on Interval Approach 2 (Two Monthly Intervals, Year One Data Only), for Different Approaches to Handling Missing Data

	Main Analysis	**Sensitivity Analysis 1[a]**	**Sensitivity Analysis 2[b]**
n	43,130	42,264	40,806
Log-likelihood	-87417	-86978	-84333
AIC	174863	173984	168694
BIC	174984	174106	168815
LRT[c]	p <0.0001	p <0.0001	p <0.0001
Average Posterior Probability (Range)	0.79-0.90	0.79-0.92	0.79-0.92
Entropy	0.72	0.72	0.73
Trajectory Class Prevalences	45.5%, 27.7%, 20.6%, 3.7%, 2.6%	45.4%, 27.8%, 20.5%, 3.6%, 2.6%	45.6%, 27.6%, 20.6%, 3.6%, 2.5%

Abbreviations: n = number of individuals; LCGA = Latent Class Growth Analysis; MSK = Musculoskeletal; AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; LRT = Likelihood Ratio Test

Individuals were excluded if an exclusion reason (age ≥67 years, death, de-registration, or last collection date) occurred:

[a] During the calendar year of the index fit note

[b] During one year follow-up post the index fit note date

[c] LRT was assessed with both the Lo-Mendell-Rubin and bootstrapped methods

Table 36. Comparison of Optimal LCGA Model Fit and Class Meaningfulness Statistics for MH Cohort, Based on Interval Approach 2 (Two Monthly Intervals, Year One Data Only), for Different Approaches to Handling Missing Data

	Main Analysis	Sensitivity Analysis 1[a]	Sensitivity Analysis 2[b]
n	62,355	60,567	57,914
Log-likelihood	-143694	-142466	-137067
AIC	287415	284960	274161
BIC	287542	285086	274287
LRT[c]	p <0.0001	p <0.0001	p <0.0001
Average Posterior Probability (Range)	0.81-0.91	0.81-0.92	0.81-0.92
Entropy	0.73	0.74	0.74
Trajectory Class Prevalences	36.5%, 32.1%, 20.4%, 5.6%, 5.5%	36.9%, 32.1%, 20.1%, 5.5%, 5.4%	36.6%, 32.0%, 20.4%, 5.6%, 5.5%

Abbreviations: n = number of individuals; LCGA = Latent Class Growth Analysis; MH = Mental Health; AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; LRT = Likelihood Ratio Test

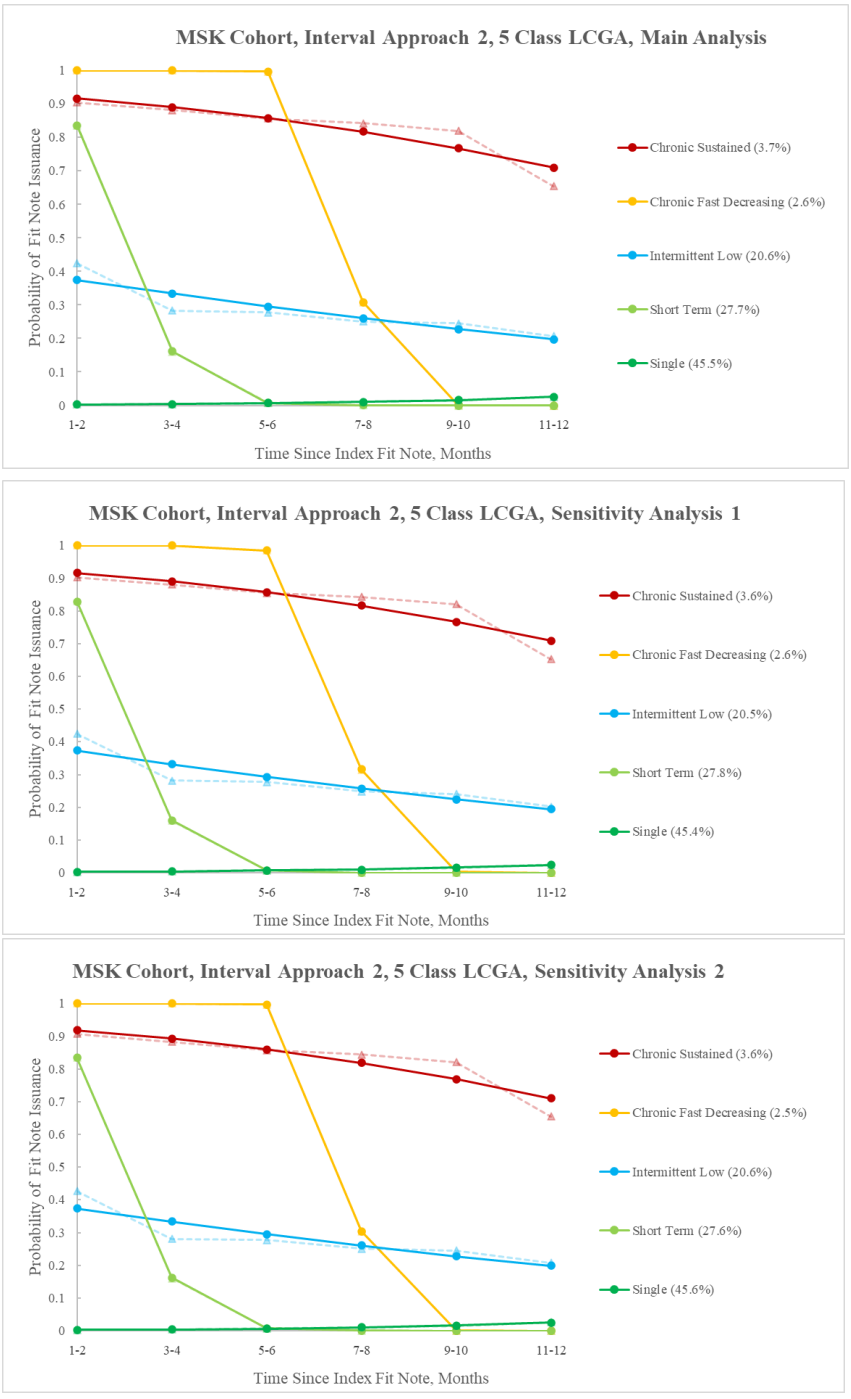
Individuals were excluded if an exclusion reason (age ≥67 years, death, de-registration, or last collection date) occurred:

[a] During the calendar year of the index fit note

[b] During one year follow-up post the index fit note date

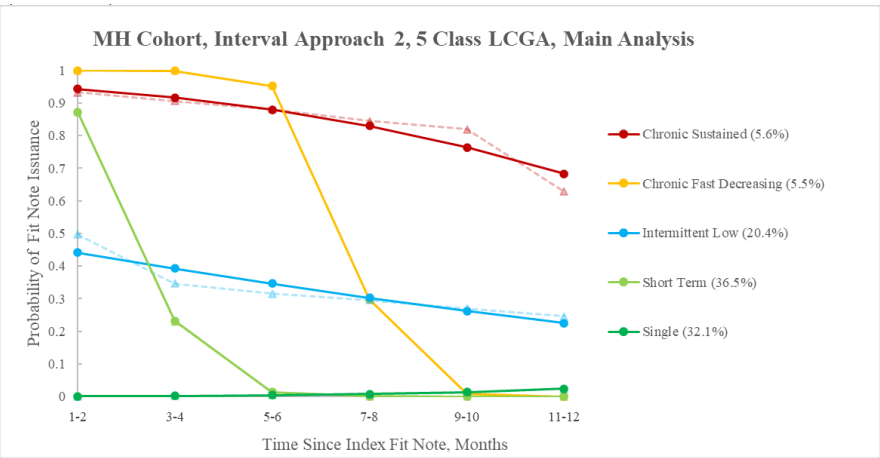
[c] LRT was assessed with both the Lo-Mendell-Rubin and bootstrapped methods

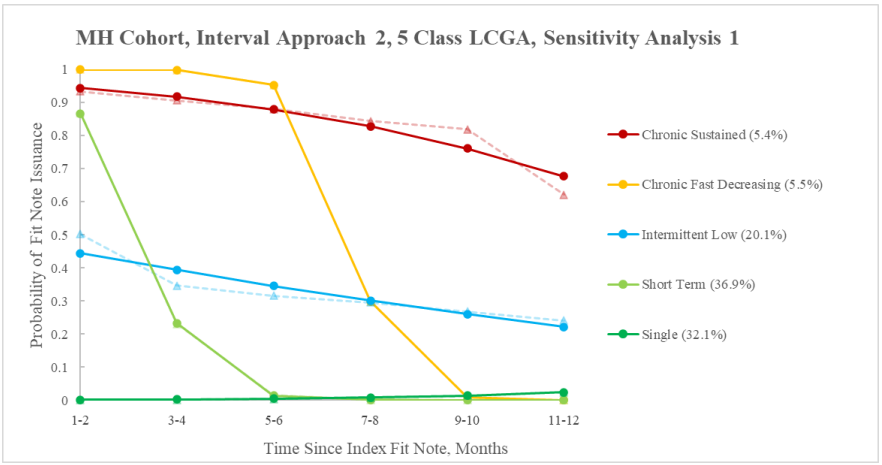
Figure 37. Comparison of Optimal LCGA Trajectory Model Plots for MSK Cohort, Based on Interval Approach 2 (Two Monthly Intervals, Year One Data Only), for: the Main Analyses (Top), Sensitivity Analysis 1 (Middle), and Sensitivity Analysis 2

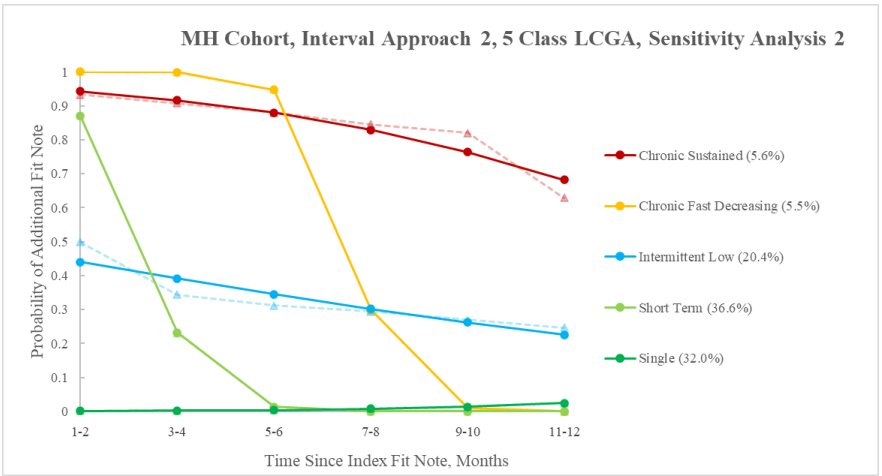


Abbreviations: LCGA = Latent Class Growth Analysis; MSK = Musculoskeletal.
For each trajectory, the solid lines represent the probability of fit note issuance in the given time interval (model estimated data), whilst the dotted lines represent the proportion of individuals issued a fit note (observed data). Trajectory class prevalence is shown, derived from a count based on posterior probability.
In sensitivity analysis 1, individuals were excluded if an exclusion reason (age ≥67 years, death, de-registration, or last collection date) occurred during the calendar year of the index fit note, and they were excluded in sensitivity analysis 2 if this occurred at any point during one year of follow-up post index fit note date.

Figure 38. Comparison of Optimal LCGA Trajectory Model Plots for MH Cohort, Based on Interval Approach 2 (Two Monthly Intervals, Year One Data Only), for: the Main Analyses (Top), Sensitivity Analysis 1 (Middle), and Sensitivity Analysis 2







Abbreviations: LCGA = Latent Class Growth Analysis; MH = Mental Health.
For each trajectory, the solid lines represent the probability of fit note issuance in the given time interval (model estimated data), whilst the dotted lines represent the proportion of individuals issued a fit note (observed data). Trajectory class prevalence is shown, derived from a count based on posterior probability.
In sensitivity analysis 1, individuals were excluded if an exclusion reason (age ≥67 years, death, de-registration, or last collection date) occurred during the calendar year of the index fit note, and they were excluded in sensitivity analysis 2 if this occurred at any point during one year of follow-up post index fit note date.

7.5.4 Summarising the Final Optimal Trajectory Models

Finally, in this Section the chosen optimal trajectory derivation models are summarised. Firstly, the nature of the shapes of the optimal derived trajectories are described. Then, the observed data patterns of individuals are explored by each of the trajectory classes, to assess within-class variability and meaningfulness of the classes. To conclude, a description of the observed follow-up fit note proportions for years two and three (unused in the model) is presented.

For both the index MSK and MH condition fit note cohorts, the optimal chosen model was the same: a five-class LCGA, under interval approach 2 (based on two-monthly recurring intervals of a one-year follow-up post index fit note date). The trajectory shapes of these five class models were similar across the two cohorts (as shown earlier in Figure 35), and consisted first of two classes characterised by low sickness absence throughout follow-up:

- 'Single', whereby the probability of fit note issuance remained close to 0 for all two-monthly intervals in the first year of follow-up (excluding the index fit note)
- 'Short Term', whereby there was an initial high probability of fit note issuance in the first two months post index fit note (of around 0.85-0.90), which then decreased sharply to under 0.25 between months three and four of follow-up, and from months five to twelve remained at close to 0.

Then there were two classes characterised by a high probability of absence:

- 'Chronic Sustained', whereby the probability of fit note issuance remained largely high (between 0.68 to 0.94) throughout all two-monthly intervals in the first year of follow-up

- ‘Chronic Fast Decreasing’, this trajectory started off similar to the ‘Chronic Sustained’ class, with a high and sustained probability (of close to 1) of fit note issuance over two-monthly intervals in the first six months of follow-up. However, from months seven to eight there was a rapid decrease to around 0.3 probability of fit note issuance, and then from months nine to twelve this probability decreased and remained close to 0.

The final derived trajectory was more difficult to categorise:

- ‘Intermittent Low’, here the probability of fit note issuance was between 0.2 and 0.45 throughout all two-monthly intervals in the first year of follow-up

The individual variability in fit note issuance over the 12 months within each of the derived classes (except the ‘Intermittent Low’ class) of the optimal trajectory models is shown in Table 37 and Table 38, for the cohort with an incident fit note due to a MSK or MH condition, respectively. These observed fit note data issuance patterns exclude the index date of the incident fit note, just as the trajectory derivation analysis did (see Section 7.2.4).

For the four trajectory classes shown, the individual variability in observed fit note issuance patterns in year one of follow-up was low, with most of the data for each trajectory class contained in a few observed fit note issuance patterns. Furthermore, the observed patterns were in line with the above descriptions of the trajectories.

For example, the ‘Single’ trajectory class was comprised largely of an observed pattern of no fit notes issued in any of the six two-monthly intervals (around 90% of the individuals in this class, across both the MSK and MH fit note analyses cohorts, followed this single pattern).

Table 37. Summary of Observed Fit Note Issuance Patterns During Year One Follow-Up for the Trajectory Classes in the Optimal Five-Class LCGA Model for the Incident MSK Condition Fit Note Cohort, Excluding the 'Intermittent Low' Trajectory Class

Trajectory Class	Outcome Pattern (2 monthly)[a]	n (%)[b]	Median Intervals With Fit Note Received
Chronic Sustained	1,1,1,1,1,1	485 (38.46%)	5
	1,1,1,1,1,0	290 (23.00%)	
	1,1,1,1,0,1	88 (6.98%)	
	1,1,1,0,1,1	80 (6.34%)	
	1,1,0,1,1,1	75 (5.95%)	
	1,1,1,0,1,0	72 (5.71%)	
	0,1,1,1,1,1	54 (4.28%)	
	1,0,1,1,1,1	50 (3.97%)	
	1,1,1,.,.,.	37 (2.93%)	
	1,1,1,1,.,.	17 (1.35%)	
Chronic Fast Decreasing	1,1,1,0,0,0	883 (66.24%)	3
	1,1,1,1,0,0	421 (31.58%)	
	1,1,1,1,0,.	11 (0.83%)	
	1,1,1,0,0,.	10 (0.75%)	
	1,1,1,0,.,.	8 (0.60%)	
Short Term	1,0,0,0,0,0	8544 (76.60%)	1
	1,1,0,0,0,0	1845 (16.54%)	
	1,.,.,.,.,.	193 (1.73%)	
	1,0,0,.,.,.	122 (1.09%)	
	1,0,.,.,.,.	116 (1.04%)	
	1,0,0,0,.,.	111 (1.00%)	
	1,0,0,0,0,.	98 (0.88%)	
	1,1,.,.,.,.	56 (0.50%)	
	1,1,0,.,.,.	30 (0.27%)	
	1,1,0,0,.,.	24 (0.22%)	
	1,1,0,0,0,.	15 (0.13%)	
Single	0,0,0,0,0,0	20117 (90.99%)	0
	0,0,0,0,0,1	737 (3.33%)	
	0,0,0,.,.,.	261 (1.18%)	
	0,0,.,.,.,.	260 (1.18%)	
	0,0,0,0,.,.	260 (1.18%)	
	0,.,.,.,.,.	256 (1.16%)	
	0,0,0,0,0,.	219 (0.99%)	

Abbreviations: LCGA = Latent Class Growth Analysis; MSK = Musculoskeletal.

[a] A value of 1 is used to denote that at least one fit note was issued in the given time interval, 0 denotes that no fit notes were issued, and a period indicates that the individual had missing data in the time interval.

[b] Cell counts less than five are not shown in accordance with CPRD reporting guidelines (to reduce risk of patient identification).
The pattern is chronologically ordered in two monthly time intervals during the first year of follow-up since index fit note. For example, a pattern of 1,0,0,0,0,0 indicates that a fit note was issued in the first two months since index fit note, but not in the ensuing 10 months.
Note: trajectory prevalence is based on most likely latent class membership, not posterior probabilities, as individuals are treated as whole persons in the observed data, hence the posterior probabilities cannot be used (as explained earlier in Section 5.2.6).

Table 38. Summary of Observed Fit Note Issuance Patterns During Year One Follow-Up for the Trajectory Classes in the Optimal Five-Class LCGA Model for the Incident MH Condition Fit Note Cohort, Excluding the 'Intermittent Low' Trajectory Class

Trajectory Class	Outcome Pattern (2 monthly)[a]	n (%)[b]	Median Intervals With Fit Note Received
Chronic Sustained (4.6%)	1,1,1,1,1,1	1061 (36.83%)	5
	1,1,1,1,1,0	734 (25.48%)	
	1,1,1,1,0,1	208 (7.22%)	
	1,1,1,0,1,1	203 (7.05%)	
	1,1,1,0,1,0	192 (6.66%)	
	1,1,0,1,1,1	152 (5.28%)	
	1,0,1,1,1,1	122 (4.23%)	
	0,1,1,1,1,1	111 (3.85%)	
	1,1,1,1,.,.	58 (2.01%)	
	1,1,1,1,1,.	22 (0.76%)	
	1,1,1,0,1,.	9 (0.31%)	
Chronic Fast Decreasing (6.2%)	1,1,1,0,0,0	2479 (64.42%)	3
	1,1,1,1,0,0	1166 (30.3%)	
	1,1,1,.,.,.	95 (2.47%)	
	1,1,1,0,0,.	48 (1.25%)	
	1,1,1,0,.,.	42 (1.09%)	
	1,1,1,1,0,.	18 (0.47%)	
Short Term (34.5%)	1,0,0,0,0,0	14824 (68.84%)	1
	1,1,0,0,0,0	4848 (22.51%)	
	1,.,.,.,.,.	479 (2.22%)	
	1,0,0,.,.,.	292 (1.36%)	
	1,0,.,.,.,.	282 (1.31%)	
	1,0,0,0,0,.	209 (0.97%)	
	1,0,0,0,.,.	195 (0.91%)	
	1,1,.,.,.,.	170 (0.79%)	
	1,1,0,0,0,.	82 (0.38%)	
	1,1,0,.,.,.	81 (0.38%)	
	1,1,0,0,.,.	72 (0.33%)	
	0,0,0,0,0,0	20669 (88.87%)	0

Single (37.3%)	0,0,0,0,0,1	733 (3.15%)
	0,.,.,.,.,.	457 (1.96%)
	0,0,.,.,.,.	386 (1.66%)
	0,0,0,.,.,.	361 (1.55%)
	0,0,0,0,0,.	328 (1.41%)
	0,0,0,0,.,.	323 (1.39%)

Abbreviations: LCGA = Latent Class Growth Analysis; MH = Mental Health.
[a] A value of 1 is used to denote that at least one fit note was issued in the given time interval, 0 denotes that no fit notes were issued, and a period indicates that the individual had missing data in the time interval.
[b] Cell counts less than five are not shown in accordance with CPRD reporting guidelines (to reduce risk of patient identification).
The pattern is chronologically ordered in two monthly time intervals during the first year of follow-up since index fit note. For example, a pattern of 1,0,0,0,0,0 indicates that a fit note was issued in the first two months since index fit note, but not in the ensuing 10 months.
Note: trajectory prevalence is based on most likely latent class membership, not posterior probabilities, as individuals are treated as whole persons in the observed data, hence the posterior probabilities cannot be used (as explained earlier in Section 5.2.6).

However, individual variability was high in the 'Intermittent Low' class. In this class there were 83 and 88 possible types of fit note issuance pattern across the MSK and MH condition fit note cohorts, respectively (shown in Appendix K).

Furthermore, the distribution of individual patterns was more dispersed in this class compared to the other four classes mentioned above. For example, in the four classes mentioned above, the majority of individuals followed one of up to five of the most commonly occurring fit note issuance patterns in the class. In contrast, in the 'Intermittent Low' class, for both the MSK and MH condition fit note cohorts, approximately 80% of individuals followed one of twenty of the most occurring fit note issuance patterns in this class.

Nonetheless, despite the individual variability in the 'Intermittent Low' class, the median number of two-monthly intervals where a fit note was issued (out of the six two-monthly intervals in the first year of follow-up data) was two intervals for both the MSK and MH condition fit note cohorts, thus these classes were indeed generally characterised by a low fit note issuance. As these fit notes could be issued at any

interval within the one year though, this more sporadic nature of this trajectory class suggested a subgroup of individuals who experienced one or more relapses of absence following an RTW during the year of follow-up.

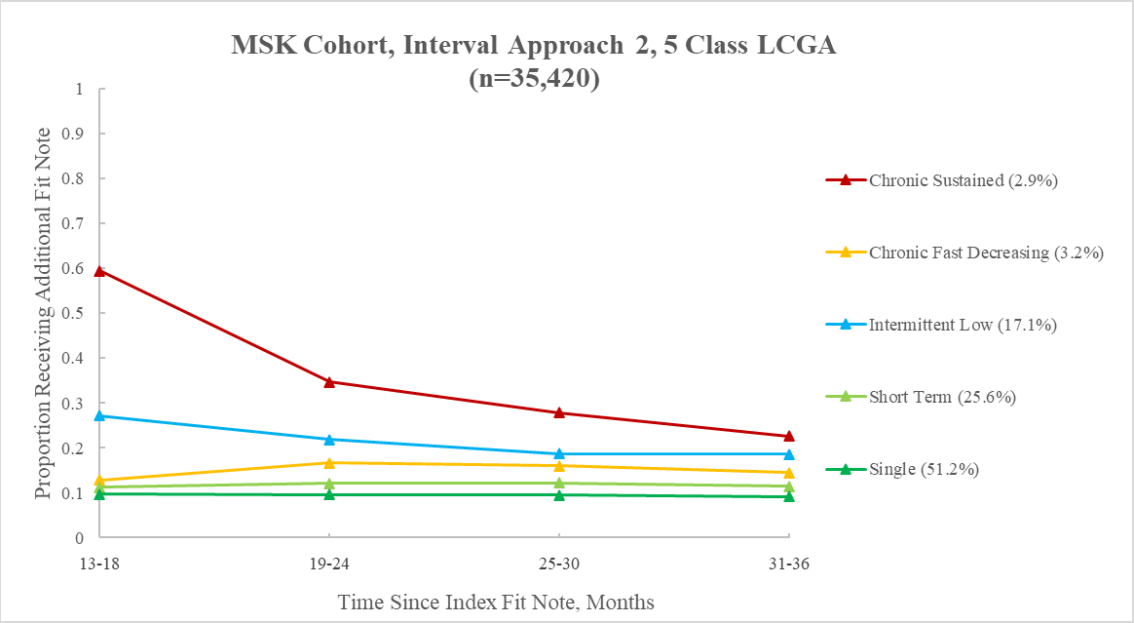
Finally, the longer-term observed fit note issuance patterns in years two and three of follow-up are presented in Figure 39, for the optimal five-class LCGA models based on incident fit notes due to a MSK or MH condition, respectively. This data is based on individuals that had complete follow-up of three-years post index fit note date.

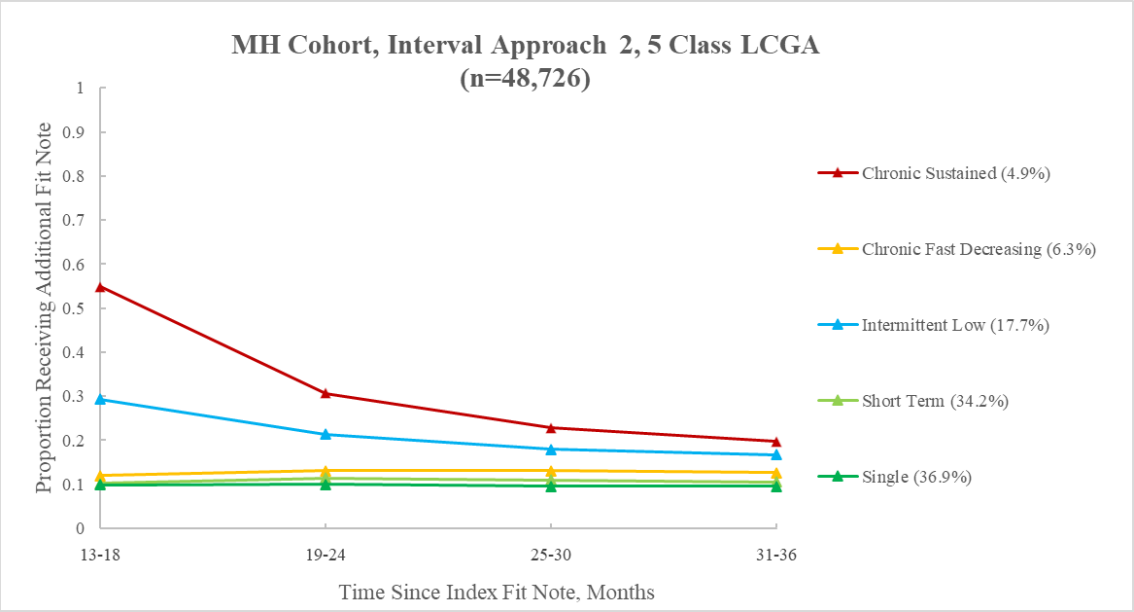
Low and stable proportions of fit note issuance were observed throughout six-monthly intervals in years two and three of follow-up in the: 'Single', 'Short Term' and 'Chronic Fast Decreasing' classes. Individuals in the 'Single' class had the lowest proportions of longer-term fit note issuance, followed by the 'Short Term' and then the 'Chronic Fast Decreasing' classes. Trends were the same for both cohorts, irrespective of whether the index fit note was due to a MSK or MH condition.

In contrast, the 'Intermittent Low' class began with a higher observed fit note issuance proportion than the above three classes, at around 30% fit note issuance in the first six months of the second year of follow-up, this then decreased steadily over the ensuing eighteen months.

The most severe trajectory class, 'Chronic Sustained', exhibited the highest proportion of fit note issuance in the first six months of the second year of follow-up (60% and 55%, for the cohort with an incident fit note due to a MSK or MH condition, respectively). Whilst this proportion decreased over the next eighteen months, it remained higher (at all time points) than observed for the other four trajectory classes.

Figure 39. Observed Fit Note Issuance Patterns in Years Two and Three of Follow-Up for MSK (Top) and MH (Bottom) Condition Optimal Five-Class LCGA Models, Under Interval Approach 2 (Two Monthly Intervals, Year One Data Only)





Abbreviations: LCGA = Latent Class Growth Analysis; MSK = Musculoskeletal; MH = Mental Health. Trajectory class prevalence is shown, derived from a count based on most likely latent class membership, not posterior probabilities, as individuals are treated as whole persons in the observed data, hence the posterior probabilities cannot be used (as explained earlier in Section 5.2.6). Furthermore, this graph is only based on individuals that had complete follow-up data for three years post index fit note date.

7.6 Discussion

7.6.1 Summary of Main Findings and Comparison with Other Research

In this study, a thorough exploration of trajectory derivation analysis using latent class methods was carried out, for two cohorts of individuals, with an incident fit note issued from 2016 to 2018 due to either a MSK or MH condition. Patterns of fit notes issued for any reason up to three-years post the index MSK or MH condition fit note date were analysed, to assess for presence of any common subgroups of individuals with similar work absence behaviour.

It was not feasible to use a continuous definition of fit notes in the trajectory derivation analysis, due to substantial missing fit note duration data. Hence a binary definition was used, for fit note issuance in a given interval. Five different approaches to defining time intervals and follow-up lengths were tested for best operational performance.

Optimal models of trajectories of absence were indeed identified, for the index MSK and MH condition fit note cohorts, which addressed the overall aim of this Chapter. These optimal models were the same for both the MSK and MH condition fit note cohorts and derived using LCGA and interval approach 2 – whereby fit note issuance was assessed through two-monthly recurring intervals over a one-year follow-up post index fit note.

There were five trajectory classes uncovered in the optimal models, two of which were more severe work absence classes, and characterised by greater issuance of fit notes: ‘Chronic Sustained’ and ‘Chronic Fast Decreasing’. The ‘Chronic Sustained’ class had a sustained high probability of fit note issuance over each two-monthly interval during the one-year follow-up period. The ‘Chronic Fast Decreasing’ class shared this pattern

of a high probability of fit note issuance for the first six months of follow-up, before a rapid decrease to a probability of 0 from months seven to ten, which was then sustained from months ten to twelve. These two trajectories of sustained absence were the least common of the five derived trajectories. Out of the individuals who had a first ever fit note due to a MSK condition, 6.3% went on to follow one of these two longer-term absence trajectories, and 11.1% for the incident MH condition fit note cohort.

Two less severe classes were also identified: 'Short Term' and 'Single', the former was characterised by a high probability of fit note issuance in the first two-months following index fit note, which then rapidly decreased to 0 by month six, and was sustained at this level until the end of one year follow-up. Whereas in the 'Single' class the probability of fit note issuance remained at close to 0 throughout each two-monthly interval during follow-up. These two classes of low work absence were the most common of the five derived trajectories, with a combined prevalence of 73.2% and 68.6% for the MSK and MH condition fit note cohorts, respectively.

The fifth class, 'Intermittent Low' (prevalence of 20.6% and 20.4% for the MSK and MH condition fit note cohorts, respectively), showed considerable heterogeneity and lack of a clearly identifiable pattern other than being episodic fit notes, suggesting a subgroup of individuals that achieved a RTW for a short time followed by one or more relapses back into absence.

Face validity of the trajectories in these optimal models was affirmed through discussion with AL's supervisory team, a GP, OHID, as well as a PPIE group. All five trajectory classes had clinical relevance and were deemed plausible in a work absence context. These chosen optimal five-class LCGA models also performed well with respect to statistical measures of model fit and separability of classes. Furthermore,

these optimal models, based on using all available data and including individuals with incomplete follow-up, were robust to two sensitivity analyses that considered other approaches to handling missing data.

As mentioned in Chapter 3, literature is scarce on trajectories of work absence, as this remains a novel area of research. Hence comparisons to our study are limited. Only six comparable trajectory derivation studies were identified in our systematic review, that contained individuals with a baseline work absence due to either a MSK or MH condition.

Nonetheless, the finding that the most severe work absence trajectories generally occurred with lower prevalence, whilst the least severe trajectories were the most prevalent, was also observed in the six studies of the systematic review.

Two of these studies, Spronken et al[73] and Farrants et al (2019),[69] also identified five trajectory classes in their final optimal model, whilst Farrants et al (2018)[72] and Rysstad et al[101] identified six classes.

Additionally, there were some similarities in the nature of the uncovered trajectory classes. Trajectories involving a sustained high level of work absence throughout follow-up, analogous to our most severe 'Chronic Sustained' class (3.7% and 5.6% prevalence, for the MSK and MH condition fit note cohorts, respectively), were identified by: Rysstad et al[101] (a 'Persistent High' class, 18.2% prevalence), Farrants et al (2018)[72] (a 'Constant High' class, 11% prevalence), Farrants et al (2019)[69] (a 'Late Decrease' class, 8.0%), and McLeod et al[70] (a 'Constant SA' class, 3.0% prevalence).

Classes of individuals who exhibited less severe, 'Short Term' behaviour in terms of their sickness absence, and had a sustained low fit note issuance after around month

four post index fit note date onwards (27.7% and 36.5% prevalence, for our MSK and MH cohorts, respectively) were also discovered by: Rysstad et al[101] (a 'Fast Decrease' class, 27.0% prevalence), Farrants et al (2018)[72] (a 'Decrease to 0 after 4 months' class, 43.0% prevalence), Farrants et al (2019)[69] (a 'Fast Decrease' class, 36.0% prevalence), McLeod et al[70] (a 'Short-delayed RTW' class, 30.6% prevalence), Pedersen et al[100] (a 'Fast RTW' class, 21.9% prevalence), and Spronken et al[73] (a 'Fast RTW without relapse' class, 60% prevalence).

7.6.1.1 Most Appropriate Time Intervals

During our optimal model selection process, five different approaches to time intervals and follow-up lengths were compared, and ultimately models based on interval approach 2 were chosen (using two-monthly recurring intervals based on a one-year follow-up post index fit note).

In particular, the trajectory derivation results based on a longer-term follow-up of three-years (approaches 3-5) did not perform as well as those based on one-year follow-up (approaches 1-2). For example, for the analysis based on repeated two-monthly intervals in year one of follow-up, when follow-up was extended from one-year (approach 2) to three years (approach 5 – with six-monthly recurring intervals for years two and three), the derived trajectories were more prone to poor visual fit.

These results suggest that it might not be possible to derive longer-term trajectories of absence. Indeed, five out of six of the included studies of trajectories of absence from our systematic review also used a one-year follow-up, in line with our optimal model choice. To extend analysis to a two or three-year follow-up might not be feasible, as an individual might have experienced different, unrelated absence episodes over such a large time span. Using a separate trajectory derivation model for different absence spells

might be a better approach. Assessing the feasibility of deriving longer-term absence trajectories remains an area for further research.

Furthermore, it was found that the shortest time interval considered in this study – repeated two-monthly intervals, outperformed longer time intervals of three- and six-months. Five out of six studies from the systematic review used monthly time intervals (and the sixth study weekly time intervals) in their optimal absence trajectory models. Time intervals under two months were not practical to test in this study, as latest NHS Digital showed 50% of MSK and MH condition fit notes have duration longer than one month.[38] Furthermore, it is difficult to compare our choice of time interval duration to other studies due to differences in absence definitions between studies.

7.6.1.2 Most Appropriate Trajectory Methodology

In this study's model building strategy, different types of LCA methods were tested for optimal performance: LGCM, LCGA, and GMM. LCGA models performed best in this study, whilst GMM models were fit unsuccessfully due to computational issues. Additional modelling for interval approach 4 based on a longer-term follow-up of three-years was also performed using piecewise modelling. Whilst using a piecewise framework did allow LCGA models to be fit under approach 4 (two-class LCGA models did not converge in a non-piecewise framework), the models under interval approach 4 were not the best performing in this study.

LCGA was also used in three other studies of sickness absence trajectories: Rysstad et al,[101] and Farrants et al in their 2018[72] and 2019[69] studies.

One key difference to the Rysstad et al[101] study in particular, was in the functional form used in the LCGA models. During our model building strategy, a linear functional form

was ascertained as most appropriate for our data. The quadratic functional form LGCM's did not converge. In contrast, the final optimal LCGA model elected by Rysstad et al,[101] featured quadratic and cubic functional forms for the included trajectory classes. Furthermore, whilst the entropy values of our optimal models (0.72 and 0.73, for the MSK and MH cohorts, respectively), were above our guideline threshold of 0.7, which indicated that there were no considerable concerns with class separability, Rysstad et al's[101] optimal model had a higher entropy of 0.95. This might be due to Rysstad et al[101] using a different method of absence measurement through a continuous definition based on number of absence days, which might allow for the identification of more nuanced patterns of absence over time than a binary definition, as in our study.

7.6.1.3 Are the Optimal Trajectories Affected by the Reason for Incident Absence?

The final objective of this study was to assess whether the derived trajectories were a function of reason for sickness absence onset. All throughout this study, when comparing like-for-like, (i.e., comparing the same interval approach and same class LCGA model), the derived trajectory classes for individuals with an index MSK condition fit note were generally similar to those for individuals with an index MH condition. Indeed, the chosen optimal LCGA models for the MSK and MH condition fit note cohorts were highly similar in the shape of the derived trajectories, as well as the statistical performance of the models. Only the prevalence of the optimal trajectories showed some level of variation; the MSK condition cohort contained slightly more individuals in the less severe absence trajectories than the MH condition cohort, whilst in the latter cohort prevalence of the most severe absence trajectories was slightly higher.

No clear differences in optimal absence trajectories were observed in this thesis' systematic review either, when comparing the three baseline MSK condition absence studies to the three baseline MH condition absence studies.

7.6.2 Study Strengths and Limitations

One of the key strengths of this study was that it was the first to explore trajectories of work absence for an English population. Furthermore, as demonstrated in Section 7.2.2, using CPRD (Aurum) data enabled our study to be highly powered for detecting trajectory prevalences. Indeed, compared to most of the studies of trajectories of sickness absence uncovered in our systematic review, our study contained a substantially greater sample size. Additionally, by using a large sample based on CPRD Aurum data, we could be confident that our cohort data was representative of the general English population.[128]

Another key strength of this study was that high quality trajectory reporting was conducted using the GRoLTS checklist,[173] and a variety of statistical measures were used to guide the choice of the optimal models, to ensure robustness of the final selection. As shown in the earlier review of sickness absence trajectory methodology in Section 5.4, there was a need to improve the general quality of trajectory reporting. Furthermore, unlike any included absence trajectory studies from the systematic review of this thesis, consideration towards other derivation methods was given in this study, specifically through using LCGA alongside attempts to fit GMM models. Attention was also shown towards selection of appropriate interval and follow-up lengths in this study through consideration of five different approaches, which is important as choice of interval and follow-length can affect the derived trajectories.[100,173]

However, one of the main limitations was that it was not possible to perform a trajectory derivation analysis based on duration of fit note. Fit note duration data, a key component for a continuous work absence definition, was largely missing in our CPRD data. If this was available, the trajectory models might have been able to uncover subgroups of individuals with even more sophisticated work absence behaviour over time. More complex trajectory functional forms would have likely been needed in this instance too.

Additionally, a 'Single with Later Relapse' trajectory class was identified in one of our LCGA models, which was of clinical interest to the GP who provided her feedback to our findings, as this could be a subgroup for HCPs to monitor closely to prevent occurrence of a transition from low to higher risk of absence over time. Perhaps if duration data were used, this subgroup of individuals who are at risk of experiencing a later episode of sickness absence, after recovering from their initial one, might have been observed in more of our models. Although, this trajectory class occurred with low prevalence (2.9%) and may have been a part of the more commonly occurring 'Intermittent Low' class that was also identified.

Finally, it must be stated that our models were based on issued fit notes and not workplace data. A recurring limitation of using CPRD data in all the studies of this thesis, is that RTW data is not available. Therefore, we have had to infer that not being issued a fit note might be indicative of a participant's RTW. If we had RTW data available, we could validate this assumption. However, it might be the case, for example, that certain subgroups such as older individuals have less fit notes issued simply because they have retired. Other individuals may have left the workforce too, for various reasons such as unemployment or becoming economically inactive (despite our

best efforts to exclude such individuals through our 16-66 years age inclusion criteria). This will be addressed in part in our next Chapter, where we investigate whether there are specific profiles of individuals (such as older individuals) that are more likely to belong to our derived trajectories.

7.6.3 Conclusion

In this first ever study of trajectories of work absence in an English population, LCGA was used to uncover five optimal trajectories over a one-year follow-up post onset of incident work absence due to either a MSK or MH condition. This included two groups of individuals with more severe fit note issuance patterns ('Chronic Sustained' and 'Chronic Fast Decreasing' trajectory classes), two less severe groups ('Single' and 'Short Term'), and a fifth group characterised by sporadic absence spells ('Intermittent Low').

The next chapter will assess health and sociodemographic characteristics associated with these optimal trajectory classes.

Chapter 8. Study 3: Assessing Association of Characteristics with Optimal Trajectories of Work Absence

8.1 Study Aim and Objectives

In the final study of this thesis, the aim was to test for presence of any associations between sociodemographic or health characteristics including types of treatment received and comorbidity with the five trajectory classes of fit note issuance derived from the chosen optimal model in the previous Chapter (for a population with an index fit note due to either a MSK or MH condition).

Specifically, the study research objectives were to answer the following questions:

1) What are the typical profiles of individuals within each of the identified work absence trajectories?

2) Is it possible to identify health and sociodemographic characteristics associated with future persistent or recurrent work absence?

3) Do the typical profiles, and any observed associations of characteristics with work absence trajectories differ by reason for index fit note?

The full set of characteristics explored in this study are detailed in the next Section (8.2), followed by a description of the analyses conducted (Section 8.3). The results of this study, for the cohort that had an index fit note due to a MSK or MH condition, are then presented in Sections 8.4 and 8.5, respectively.

8.2 Defining the Characteristics Explored

The definitions of the characteristics used in this study are provided in this Section, alongside details of any data manipulation that was performed pre-analysis, such as categorising characteristics into a smaller number of groups to remove any sparse categories.

8.2.1 Sociodemographic Characteristics

Four sociodemographic characteristics were explored: sex, age, geographical region, and deprivation. The first three characteristics were retained as originally defined and used in Section 6.2.6 from study 1, except that geographical region was condensed from nine categories into three:

- North of England: Northeast, Northwest, Yorkshire and the Humber
- Middle of England: East Midlands, West Midlands, East of England
- South of England: Southeast, Southwest, London

This was done to eliminate sparse categories, and to facilitate interpretation of the results. For example, for the cohort with an index MH condition fit note, the following regions each contained <5% of the total sample data: Northwest, Yorkshire and the Humber, East Midlands and East of England.

The fourth sociodemographic characteristic, deprivation, was defined through the commonly used English Index of Multiple Deprivation (IMD) measure,[200] which was accessed through a data-linkage request in the CPRD RDG protocol (Appendix F). The latest version of linked IMD data available was used (the 2019 version). This is a relative measure, rather than absolute, that compares levels of deprivation of English neighbourhoods at a lower-layer super output area (LSOA) level.

Other measures of deprivation are available through linkage with CPRD, such as the Townsend Score or Carstairs Index.[200] However, the latter two measures are more focused on material deprivation, whereas the IMD was chosen as it covers a broader range of domains in its definition of deprivation. Nonetheless, all three of these deprivation measures have been shown to be highly correlated, thus the choice of one over the other is not expected to have a substantial impact on analyses.[200]

CPRD linked IMD data is available at quintile, decile, or 'twentile' format. Quintiles were chosen to keep the number of deprivation categories at a manageable level. Thus, IMD scores range from 1 (least deprived) to 5 (most deprived) in this study.

For age there was no missing data. For sex and region, individuals with an 'indeterminate' coded sex or no region data, respectively, were excluded due to low counts (consistent with the approach taken in Sections 6.2.6 and 6.2.7 in study 1).

For IMD, and indeed any other new characteristics in this study not previously used in this thesis, the approach to any missing data was assessed on a case-by-case basis for each characteristic. IMD data was missing for 3-4% of the sample and this data was retained as a 'missing' IMD category.

8.2.2 Health Characteristics

In total, six health characteristics were explored in this study. Two such characteristics concerned prior and follow-up consultation patterns for MSK and MH conditions.

Prior consultation patterns, for the MSK cohort, were defined as the number of MSK consultations in the two years prior to index fit note date (excluding the index MSK consultation). Similarly, for the MH cohort, number of MH consultations in the prior two years were explored (excluding index MH consultation).

Only MSK or MH consultations unique by consultation date were counted (for example, if a participant had two MSK consultations recorded on the same date within this time period, this was counted as a '1', not '2'). MSK and MH consultations were defined using the same code lists as mentioned in Section 6.2.2 of study 1.

The definition of follow-up patterns of MSK/MH consultations was similar to the prior patterns characteristic, except that consultations were now counted if they occurred during a three-year follow-up period post index fit note date (excluding the index MSK/MH consultation again).

If no recorded MSK or MH consultations were found in the CPRD Aurum database for an individual during either of these two time periods (prior to or during follow-up), it was assumed that the individual had no prior or follow-up MSK/MH consultations.

After extracting the data and exploring the distributions of these prior and follow-up consultation patterns, it was decided by AL and his supervision team to apply the following categorisation in analyses: 0, 1, 2 or ≥3 consultations.

Next, the type of MSK or MH condition that was linked to the incident fit note was also explored.

For the index MSK condition fit note cohort, this characteristic explored which of the six MSK conditions the index fit note was issued for (osteoarthritis, inflammatory MSK, back pain, knee pain, hip pain, hand/wrist pain – as in Section 6.2.2).

During the process of extracting the data for this index MSK consultation (i.e., the most recent MSK consultation that was recorded in the two weeks prior to or on the index fit note date), some duplicates entries (0.5% of the index MSK cohort) were discovered where an individual had more than one type of MSK consultation recorded on this most

recent MSK consultation date. De-duplication was performed by creating a hierarchy based on occurrence of each of these six types of MSK condition in the dataset, whereby MSK condition types with greater occurrence were given higher priority.

Then, for the type of index MH condition, MH conditions were originally defined using n=189 codes relating to either: stress, anxiety, and/or depression (Section 6.2.2). However, in the original code list used, it was not specified which of these three MH conditions the n=189 codes related to. Therefore, AL first independently categorised these n=189 codes, and then finalised this categorisation following feedback from GWJ, into: stress, anxiety, depression, or anxiety and depression (some Read/SNOMED codes contained reference to both conditions). The categorised MH condition code list is provided in Appendix L.

Duplicate types of index MH condition occurred for approximately 2% of the index MH condition cohort. However, as occurrence of each of the four MH conditions were similar (around 25%), a different approach to the type of index MSK condition was taken for de-duplication. Firstly, any of the three MH conditions relating to anxiety, depression, or anxiety and depression combined, were prioritised over stress. For example, if a participant had consultations for stress and anxiety both recorded on the same index date, only the anxiety consultation was retained. Then, for remaining duplicates, only ‘anxiety and depression’ consultations were retained, to cover all remaining possibilities. For example, if a participant had a consultation for anxiety, and another for anxiety and depression, both recorded on the same date, only the latter data was preserved.

Next, a health characteristic for smoking status was included in this study. Smoking status was identified from a Read/SNOMED code list (n=361) from the MSKCOM

study[125] (code list publicly available from https://doi.org/10.21252/878s-x990). After removing sparsely populated smoking categories, smoking codes were ultimately categorised into the following types of smoking status in this study: current, ex-smoker, never, and not recorded (i.e. individuals with missing smoking data).

To create these categories, smoking codes were searched for in the CPRD Aurum database five years prior to (and including) index fit note date, and the most recent smoking status data retained. Five years was chosen in the search to allow a broad timeframe, as it was anticipated that there would be a higher level of missing data for this characteristic (informed by prior experience from other Keele-based CPRD researchers). Indeed, no smoking status records were found in these five years for 14.3% of each of the index MSK and MH condition fit note cohorts.

Multiple imputation of missing smoking data was considered, but it was unlikely that the required Missing at Random (MAR) assumption would hold for such analyses to be valid. This is because smoking data is not recorded routinely, but only upon request from the HCP, which is not believed to occur at random. For example, recording of smoking data may be influenced by the participant being diagnosed for a disease such as Chronic Obstructive Pulmonary Disease (COPD), as smoking is known to be strongly associated with COPD.[201]

The penultimate health characteristic explored in this study, was body mass index (BMI). Recorded BMI data was also identified from a Read/SNOMED code list (n=233) from the MSKCOM study[125] (code list publicly available from https://doi.org/10.21252/878s-x990). This was the most complex characteristic derived in this study and required a multi-faceted approach (in line with other Keele-based CPRD studies) – this is provided in full in Appendix M.

In brief, BMI was calculated using the most recent BMI code or value (or height and weight if BMI value was not recorded). Then, BMI was ultimately categorised in accordance with NICE guidelines[202] into:

- Underweight/Normal ($10 \leq BMI < 25$)
- Overweight ($25 \leq BMI < 30$)
- Obese ($30 \leq BMI < 80$)
- Not Recorded (no BMI data available, or BMI outside of a plausible range, defined as: $BMI < 10$ or $BMI \geq 80$)

Similar to smoking status, BMI codes were also searched for in the five years prior to (and including) index fit note date, as a relatively high level of missing BMI data was also anticipated. The BMI characteristic did ultimately have a substantial amount of missing data, with 35.8% and 37.0% 'not recorded' BMI data, for the index MSK condition fit note and index MH condition fit note cohorts, respectively. This was the highest level of missing data out of all the characteristics considered in this study. Multiple imputation was considered, but as with smoking, not feasible due to the MAR assumption unlikely to hold true.

The final health characteristic considered in this study, was called a 'contact count'. This was intended to be a count of consultations over the two years prior to index fit note date, and such data was searched for in the Observation CPRD Aurum files. Specifically, all records in the two years prior to index fit note were searched and summarised by a count of unique date entries per participant.

However, the Observation files in CPRD Aurum contain not only medical consultation data, rather, full medical history data.[203] This includes clinical measurements (for

example, recording of smoking and BMI data), as well as laboratory test results, and symptoms.[203] Therefore, each participant can potentially have a vast amount of entries in these Observation files over a short time period.

For example, each time a height entry is recorded, it appears in the files as a separate data row. Or if a new smoking record is taken, a new data row would again appear. To isolate only medical consultation data in these Observation files is challenging, as it would require defining code lists specific to each medical condition of interest. Comorbidity is covered in Section 8.2.4, therefore the approach taken in this study was to treat this characteristic in more of an exploratory manner and use it for descriptive analyses.

For the index MSK condition fit note cohort, all MSK consultations were excluded from this contact count, and MH consultations were excluded for the index MH condition fit note cohort. This was to prevent double counting, as MSK and MH consultations had already been accounted for by our prior consultation pattern characteristic.

When the extracted contact count data was investigated, there was indeed a wide range of counts present. It was decided (through discussion between AL and supervision team), to categorise as: 1-10, 11-15, 16-25, and ≥26.

8.2.3 Types of Treatment Received

In this Section, four separate characteristics relating to types of treatment received are defined: opioids, NSAIDs, gabapentinoids, and antidepressants. For consistency, all four of these treatment characteristics were used in the analysis for both the index MSK and MH condition fit note cohorts.

All four of these treatments were searched for (separately) in the Drug Issue files of CPRD Aurum,[203] from the two years prior to index fit note date and up to, but not including the index MSK consultation or MH consultation date, for the index MSK condition or MH condition fit note cohorts, respectively. Then a binary indicator variable was created for prescription of each of the four treatments (yes, if at least one treatment prescription was found in this time period, or no if not).

Opioids, NSAIDs, and gabapentinoids, were searched for using code lists from the MSKCOM study[125] (code lists publicly available from https://doi.org/10.21252/878s-x990). Specifically, these particular code lists were developed using the work of Bedson et al,[204] which categorised over 300 types of analgesia into the following:

- Group 1: Basic analgesics (this consists of paracetamol, Ibuprofen (200-400mg), Aspirin (600mg), Capsaicin, and Topical NSAIDs)
- Group 2: Weak analgesics (n=10 weak combination opioids)
- Group 3: Moderate analgesics (n=14 moderate combination opioids + opioids)
- Group 4: Strong or Very Strong analgesics (n=57 strong combination opioids + opioids, Morphine and Oxycodone))
- Group 5: NSAIDs (includes Ibuprofen at 600mg strength, as well as COX 2)
- Group 6: Gabapentinoids

Opioids were defined by combining groups 2 to 4 (weak, moderate, strong or very strong analgesics) from Bedson et al's hierarchy,[204] whilst NSAIDs were defined using Group 5, and gabapentinoids using Group 6.

Antidepressants were searched for using a code list (n=268 codes) from the MEDDIP study[124] (code list publicly available from https://www.keele.ac.uk/mrr/codelists/otherdefinitions/).

8.2.4 Comorbidity

Then comorbidity was first assessed by a polypharmacy characteristic. This was defined as a drug count in the two years prior to index fit note date and up to, but not including the index MSK consultation or MH consultation date, for the index MSK condition or MH condition fit note cohorts, respectively. Furthermore, the treatments considered thus far in the previous Section (opioids, NSAIDs, gabapentinoids, and antidepressants) were excluded from this count to avoid double counting.

It was assumed that if no drug prescriptions were found for a particular individual in this time period (excluding the four drugs mentioned), then their polypharmacy was zero. Upon examining the data, polypharmacy was categorised as: 0, 1-4, 5-9, ≥10. The definition of polypharmacy in the literature is variable, however, these categories include the commonly used definition of polypharmacy as ≥5 drugs, and further, the ≥10 category covers individuals defined as having 'hyperpolypharmacy' or 'excessive polypharmacy'.[205]

Then, the final characteristic tested in this study, was the Charlson Comorbidity Index (CCI).[206] This tool was created in 1987 as a weighted index for predicting one-year mortality rates for patients with specific comorbidities. The original CCI contained 19 comorbidities, whereby each condition was assigned a score ranging from 1 to 6, based on weights derived from one-year mortality hazard ratios using a Cox proportional

hazards model. The scores of these comorbidities could then be totalled, to give each participant a CCI score.[206]

Since its inception, the CCI has been widely used and is considered a gold standard for assessing comorbidity in clinical research, and has undergone various modifications for use with different data sources.[207] The version of CCI used in this study, was based on a code list (n=9949) from the MSKCOM study[125] (code list publicly available from https://doi.org/10.21252/878s-x990). This included n=17 conditions, as shown in Box 5, alongside the assigned condition scores.

Box 5. Application of Charlson Comorbidity Index in this Study

Condition - Score
AIDS - 4
Congestive Heart Failure (CHF) - 1
Cancer - 2
Chronic Obstructive Pulmonary Disease (COPD) - 1
Dementia - 1
Diabetes (Uncomplicated) - 1
Diabetes (Complicated) - 2
Hemiplegia - 2
Metastatic Cancer - 4
Myocardial Infarction (MI) - 1
Mild Liver Disease - 1
Moderate Liver Disease - 3
Peptic Ulcer Disease - 1
Peripheral Vascular Disease (PVD) - 1
Renal Disease - 2
Rheumatic Disease[a] - 1
Stroke - 1

[a] Rheumatic Disease was excluded from the Index Musculoskeletal Condition Fit Note Cohort

These 17 conditions were searched for in the two years prior to and including index fit note date, using the CPRD Aurum Observation files. Then, for any consultation data that was found relating to these conditions, the data was de-duplicated to just retain one consultation record for each condition, irrespective of the frequency of consultations.

Finally, for each participant, the number of distinct conditions that they had a consultation for were summed together using the appropriate weights (as in Box 5), to produce a final CCI score. AL and his supervisory team then decided to categorise this score, for final use in this study as: 0, 1, ≥2.

For the index MSK condition fit note cohort only, a modified CCI score was used. This involved the same n=17 conditions as in Box 5, except that rheumatic disease was excluded (to avoid double counting this MSK condition).

It was assumed that if consultations relating to these n=17 conditions were not found in the Observation files, then the participant did not have such consultations during this time period.

8.2.5 Summary

In total 16 characteristics were considered in this study:

- Sociodemographic Characteristics (n=4)
- Health Characteristics (n=6)
- Types of Treatment Received (n=4)
- Comorbidity (n=2)

These characteristics are summarised in Table 39.

Table 39. Summary of the n=16 Characteristics Explored in this Study

Type of Characteristic	Characteristic	Category Definitions	Cohort	
			MSK Index Fit Note	MH Index Fit Note
Sociodemographic	Sex	Female, Male	✓	✓
	Age	16-25, 26-35, 36-45, 46-55, 56-65 (years)	✓	✓
	Region[a]	North of England, Middle of England, South of England	✓	✓
	IMD[b]	1-5, Missing	✓	✓
Health	MSK Consultations - Prior 2 Years[c]	0, 1, 2, ≥3	✓	×
	MSK Consultations - 3 Year Follow-up[c]	0, 1, 2, ≥3	✓	×
	MH Consultations - Prior 2 Years[d]	0, 1, 2, ≥3	×	✓
	MH Consultations - 3 Year Follow-up[d]	0, 1, 2, ≥3	×	✓
	Baseline MSK Condition	Back pain, Knee pain, Hand/Wrist pain, Hip pain, Inflammatory MSK, Osteoarthritis	✓	×
	Baseline MH Condition	Stress, Anxiety, Depression, Anxiety and Depression	×	✓
	Smoking Status	Never, Current, Ex Smoker, Not Recorded	✓	✓
	BMI[e]	Underweight/Normal, Overweight, Obese, Not Recorded	✓	✓
	Contact Count[f]	1-10, 11-15, 16-25, ≥26	✓	✓
Types of Treatment Received	Opioids	Yes, No	✓	✓
	NSAIDs	Yes, No	✓	✓
	Gabapentinoids	Yes, No	✓	✓
	Antidepressants	Yes, No	✓	✓

Comorbidity	Polypharmacy[g]	0, 1-4, 5-9, ≥10	✓	✓
	CCI Score	0, 1, ≥2	×	✓
	Modified CCI Score[h]	0, 1, ≥2	✓	×

Abbreviations: IMD = Index of Multiple Deprivation; MSK = Musculoskeletal; MH = Mental Health; NSAIDs = Non-Steroidal Anti-Inflammatory Drugs; BMI = Body Mass Index; CCI = Charlson Comorbidity Index

[a] North of England defined as: North East, North West, Yorkshire and the Humber; Middle of England: East Midlands, West Midlands, East of England; and South of England: South East, South West, London

[b] Quintiles are used for IMD (1-5), where a higher score represents more deprived areas

[c] Excluding the index MSK consultation

[d] Excluding the index MH consultation

[e] Underweight/Normal: 10<=BMI<25; Overweight: 25<=BMI<30; Obese: 30<=BMI<80; Not Recorded: no BMI data available or BMI<10 or BMI>=80

[f] Count of all medical consultations (excluding MSK related for MSK cohort, and excluding MH related for MH cohort), as well as any recording of data such as BMI, smoking etc.

[g] Excluding Opioids, NSAIDs, Gabapentinoids, Antidepressants

[h] Excluding Rheumatic Disease

Two other comorbidities were also initially considered as characteristics in this study. These were the top two reasons for the highest sickness absence rates (excluding MSK and MH conditions): injury/poisoning and respiratory conditions, as identified by Wynne-Jones et al (2009).[57] However, these were later removed in the CPRD RDG protocol amendment (Appendix F), as there was a risk of double counting given the other comorbidity covariates considered. For example, the definition of CCI used in this study includes COPD, which overlaps with respiratory conditions.

8.3 Analysis Plan

The analysis approach used in this study was three-fold:

- Stage 1: Descriptive Summary of Characteristics
- Stage 2: Univariable Association Analysis
- Stage 3: Multivariable Association Analysis

Each stage was repeated separately, for the index MSK and MH condition fit note cohorts.

For Stage 1, a descriptive summary of all n=16 characteristics described in Section 8.2 was performed, stratified by trajectory class. As all of the characteristics were ultimately defined in a categorical format (Section 8.2), counts and percentages were reported (at a total level for the cohort, and row percentages across the trajectory classes). This provided an indication of the profile of individuals within each of the identified work absence trajectories (to address the first and third objectives of this study).

Stages 2 and 3 were then performed to address the second and third objectives of this study and test association of characteristics with future persistent or recurrent work absence (the two most severe of the five derived trajectory classes). This was achieved by performing multinomial logistic regression.

Multinomial logistic regression is an analysis technique that allows trajectory classes to be input as a categorical outcome (on a nominal scale) in a logistic regression model, and then covariates can be added to this model as predictors, to test their influence on the outcome (i.e., the likelihood of belonging to a particular trajectory class).[141] There are two ways that multinomial logistic regression can be performed to test the effect of predictors on derived trajectory classes: using a one-step or a three-step approach.

The one-step approach allows a trajectory derivation model to be run in the same step as the multinomial regression analysis. However, the resulting derived trajectories may be different than those from an unconditional trajectory derivation model, as the trajectory classes are re-estimated with this approach, conditional on the predictors included.

For example, if a one-step approach to multinomial logistic regression was applied to the optimal five-class LCGA model from Chapter 7, the optimal five classes might change in shape and prevalence after they have been re-estimated conditional on

predictors. Thus, the trajectory-covariate associations from the one-step approach may be referring to wholly different trajectory classes to those originally derived in the optimal unconditional LCGA model. Therefore, the one-step approach was not appropriate for this study.

In contrast, the three-step approach, developed by Vermunt (2010),[208] allows the original derived trajectory classes to be preserved, and the uncertainty in class allocation is incorporated in the multinomial logistic regression analysis given individuals are allocated to the class for which they have the highest probability. This is achieved through specifying uncertainty rates associated with the derived classes (also known as misclassification error rates), which are based on the posterior probabilities. If a trajectory derivation model has low entropy, class separation will be poor, and the uncertainty error rates will be higher.

Therefore, with this three-step approach, covariates are tested for association with the same derived trajectories (i.e., with the same shape and prevalence) as in the unconditional mixture model, which is what is required in this study.

The three-step approach, was applied in the following manner:

1) The unconditional mixture model was first estimated (this was the optimal five-class LCGA model under interval approach 2, and this part of the process was performed in Chapter 7). Participant ID was retained in the output file, along with the trajectory class that each participant was assigned to, based on most likely class membership (i.e., this corresponds to the 'Final Assigned Class' column from the earlier example in Table 10 of Chapter 5). Misclassification error rates relating to the derived classes were also stored.

2) Covariate data was then added to this output file (using the characteristics described in Section 8.2).
3) Multinomial logistic regression analysis was then performed, accounting for the uncertainty in class membership to preserve the original five optimal classes. Clustering of individuals by general practice was also accounted for in the regression by use of a sampling weight method for modelling latent variables, whereby parameters were estimated by maximising a weighted log likelihood function.[209]

ORs of association are reported for model results, alongside 95% CIs. Statistically significant ORs are indicated in bold text in the Results Tables and determined by 95% CIs that do not include the OR value of 1. All analyses were run with the same number of random sets of starting values (500) and iterations (100) as in study 2, and all models converged without issue.

In stage 2 of the analyses, only univariable multinomial logistic regression models were run, i.e., characteristics were each tested separately for association with the optimal trajectory classes to provide unadjusted ORs. All n=16 characteristics described in this study were tested for (unadjusted) association, except for the pattern of MSK or MH consultations during follow-up. This characteristic was only used for descriptive analysis in stage 1, as it was illogical to include a participant's follow-up behaviour as a predictor of future work absence outcomes.

Then, in stage 3 of the analyses, multivariable multinomial logistic regression was performed. Here, association of multiple characteristics with the optimal trajectory classes was simultaneously tested, with all characteristics mutually adjusted for (thus providing adjusted ORs).

The results of the stage 2 univariable analyses were not used to inform the choice of candidate predictors for the multivariable analyses in stage 3. It is known that covariates which are statistically significant in a univariable logistic regression analysis may not be significant in multivariable analyses, and vice versa.[210] Rather, the stage 2 univariable analyses were performed as an exploratory step, to explore any differences and similarities in significant predictors with the analyses from stage 3 and assess the impact of confounding.

In stage 3, all of the remaining n=15 characteristics were originally to be included in the model. However, due to concerns with the contact count characteristic (as mentioned in Section 8.2.2), and expected multicollinearity with polypharmacy, correlation between these two characteristics was assessed. The Pearson Product-Moment Correlation Coefficient was used (this is a measure with a range of -1 to +1, with values closer to either -1 or +1 indicative of higher correlation between the two characteristics being tested), alongside a cross-tabulation of counts. If there was deemed to be sufficient presence of multicollinearity, only polypharmacy was retained as a candidate predictor, and contact count was not included in the final adjusted model. Otherwise, both of these characteristics were to be retained in the multivariable analyses.

The least severe of the five optimal work absence trajectories, ‘Single’ (characterised by a constant low probability close to 0 of fit note issuance throughout all of the one-year follow-up period post incident fit note), was chosen as the reference in all multinomial logistic regression analyses. The main comparisons of interest were against the most severe trajectory classes, ‘Chronic Sustained’, and ‘Chronic Fast Decreasing’ (as compared to the ‘Single’ class), as this is what the second study objective concerned.

Stata MP version 17.0 was used to derive and clean the characteristic data into an appropriate format. Then, this characteristic data was added to the optimal five-class LCGA trajectory output file (to create an input file for multinomial logistic regression), and descriptive analyses were also performed using this statistical software.

All multinomial logistic regression analyses were performed using Mplus Version 8.9. By default, when multinomial logistic regression is conducted in Mplus, listwise deletion is performed for any covariates with missing data. However, approaches to handle any missing characteristic data had already been applied in Section 8.2, therefore all data was retained during application of multinomial logistic regression. Furthermore, to assess for the effects of different approaches to handling individuals with missing data during the one-year follow-up period post incident fit note, i.e. the period in which the optimal trajectories were derived, the same two sensitivity analyses from Chapter 7 (see Section 7.2.8 for further details) were performed in this study too and the trajectory-covariate association results compared against those of the main analyses.

The results from stages 1 to 3 are presented first for the cohort with an index MSK condition fit note (in Section 8.4), then for the index MH condition fit note cohort (Section 8.5).

8.4 Results: Characteristics Associated with Optimal Trajectories of Work Absence Due to a MSK Condition

8.4.1 Descriptive Summary of Optimal Trajectories and Univariable Associations

As shown in the descriptive statistics in Table 40, the profiles of the two most severe work absence trajectories (‘Chronic Sustained’ and ‘Chronic Fast Decreasing’) were highly similar, for the cohort with an index MSK condition fit note.

In particular, the profiles of these two trajectory classes consisted of relatively greater numbers of individuals:

- That were older (46-66 years)
- Living in the North of England
- Living in the most deprived areas of England (IMD = 5)
- With more (≥2) MSK consultations in the two years prior to index fit note
- With more (≥3) MSK consultations during the three-year follow-up
- With baseline inflammatory MSK, osteoarthritis, or hip pain
- That were prescribed opioids, NSAIDs, gabapentinoids, or antidepressants in the two years prior to index fit note
- With a higher polypharmacy (≥5 drugs prescribed in the two years prior to index fit note, excluding opioids, NSAIDs, gabapentinoids, or antidepressants)
- That were current or ex-smokers
- That were obese
- With at least one other comorbidity (in addition to a MSK condition) in the two years prior to index fit note date (modified CCI score of ≥1)
- With a greater contact count (≥26) in the two years prior to index fit note date

The few differences between these two class profiles were that the ‘Chronic Sustained’ class consisted of relatively more individuals living in the Midlands, whilst the ‘Chronic Fast Decreasing’ class had relatively more females and individuals with baseline knee pain.

The profile of the ‘Intermittent Low’ Class was also very similar to these two most severe work absence classes.

In contrast, the profile for the ‘Single’ class was generally the opposite. For example, in this class, there was a relatively higher proportion of males, younger individuals (16-25 years), and individuals living in the South and least deprived areas of England.

Finally, the ‘Short Term’ class was harder to summarise, as the profile was largely a mix of the ‘Single’ and most severe class profiles.

To conclude this Section, the unadjusted OR results of associations of individual characteristics and the derived optimal trajectories for the index MSK condition fit note cohort are shown in Appendix N, Table N.1. Unadjusted associations of the two most severe trajectory classes, compared to the reference ‘Single’ trajectory, generally affirmed the combined descriptive profile of the ‘Chronic Sustained’ and ‘Chronic Fast Decreasing’ classes mentioned above.

Table 40. Descriptive Statistics of Optimal Trajectories of Work Absence Due to a MSK Condition

	Overall	Trajectory Class				
		Chronic Sustained	Chronic Fast Decreasing	Intermittent Low	Short Term	Single
	n=43,130	1261 (2.92%)	1333 (3.09%)	7272 (16.86%)	11154 (25.86%)	22110 (51.26%)
Female	19637 (45.53%)	577 (2.94%)	662 (3.37%)	3592 (18.29%)	5133 (26.14%)	9673 (49.26%)
Age						
16-25 years	6897 (15.99%)	111 (1.61%)	162 (2.35%)	1192 (17.28%)	1567 (22.72%)	3865 (56.04%)
26-35 years	9073 (21.04%)	189 (2.08%)	215 (2.37%)	1536 (16.93%)	2367 (26.09%)	4766 (52.53%)
36-45 years	9477 (21.97%)	253 (2.67%)	259 (2.73%)	1510 (15.93%)	2396 (25.28%)	5059 (53.38%)
46-55 years	10503 (24.35%)	350 (3.33%)	342 (3.26%)	1689 (16.08%)	2881 (27.43%)	5241 (49.90%)
56-66 years	7180 (16.65%)	358 (4.99%)	355 (4.94%)	1345 (18.73%)	1943 (27.06%)	3179 (44.28%)
Region[a]						
North of England	10475 (24.29%)	357 (3.41%)	424 (4.05%)	1733 (16.54%)	2948 (28.14%)	5013 (47.86%)
Middle of England	10185 (23.61%)	341 (3.35%)	314 (3.08%)	1781 (17.49%)	2691 (26.42%)	5058 (49.66%)
South of England	22470 (52.10%)	563 (2.51%)	595 (2.65%)	3758 (16.72%)	5515 (24.54%)	12039 (53.58%)
IMD[b]						
1	6008 (13.93%)	123 (2.05%)	154 (2.56%)	816 (13.58%)	1636 (27.23%)	3279 (54.58%)
2	7003 (16.24%)	150 (2.14%)	207 (2.96%)	1032 (14.74%)	1904 (27.19%)	3710 (52.98%)
3	7780 (18.04%)	187 (2.40%)	226 (2.90%)	1264 (16.25%)	2035 (26.16%)	4068 (52.29%)
4	9763 (22.64%)	288 (2.95%)	292 (2.99%)	1802 (18.46%)	2429 (24.88%)	4952 (50.72%)
5	11018 (25.55%)	478 (4.34%)	435 (3.95%)	2178 (19.77%)	2645 (24.01%)	5282 (47.94%)
Missing	1558 (3.61%)	35 (2.25%)	19 (1.22%)	180 (11.55%)	505 (32.41%)	819 (52.57%)
MSK Consultations - Prior 2 Years[c]						

0	21642 (50.18%)	456 (2.11%)	530 (2.45%)	3292 (15.21%)	5407 (24.98%)	11957 (55.25%)
1	11135 (25.82%)	331 (2.97%)	323 (2.90%)	1828 (16.42%)	2957 (26.56%)	5696 (51.15%)
2	4705 (10.91%)	184 (3.91%)	211 (4.48%)	892 (18.96%)	1271 (27.01%)	2147 (45.63%)
≥3	5648 (13.10%)	290 (5.13%)	269 (4.76%)	1260 (22.31%)	1519 (26.89%)	2310 (40.90%)
MSK Consultations - 3 Year Follow-up[c]						
0	16804 (38.96%)	123 (0.73%)	179 (1.07%)	1844 (10.97%)	2310 (13.75%)	12348 (73.48%)
1	9118 (21.14%)	104 (1.14%)	156 (1.71%)	1400 (15.35%)	2901 (31.82%)	4557 (49.98%)
2	5382 (12.48%)	104 (1.93%)	155 (2.88%)	1039 (19.31%)	1974 (36.68%)	2110 (39.20%)
≥3	11826 (27.42%)	930 (7.86%)	843 (7.13%)	2989 (25.27%)	3969 (33.56%)	3095 (26.17%)
Baseline MSK Condition						
Back pain	30232 (70.10%)	757 (2.50%)	833 (2.76%)	4783 (15.82%)	8051 (26.63%)	15808 (52.29%)
Knee pain	6424 (14.89%)	194 (3.02%)	228 (3.55%)	1180 (18.37%)	1578 (24.56%)	3244 (50.50%)
Hand/wrist pain	2127 (4.93%)	53 (2.49%)	68 (3.20%)	356 (16.74%)	483 (22.71%)	1167 (54.87%)
Inflammatory MSK	1532 (3.55%)	61 (3.98%)	53 (3.46%)	291 (18.99%)	362 (23.63%)	765 (49.93%)
Osteoarthritis	1509 (3.50%)	120 (7.95%)	94 (6.23%)	404 (26.77%)	334 (22.13%)	557 (36.91%)
Hip pain	1306 (3.03%)	76 (5.82%)	57 (4.36%)	258 (19.75%)	346 (26.49%)	569 (43.57%)
Opioids	13117 (30.41%)	585 (4.46%)	596 (4.54%)	2640 (20.13%)	3522 (26.85%)	5774 (44.02%)
NSAIDs	13256 (30.73%)	511 (3.85%)	503 (3.79%)	2561 (19.32%)	3522 (26.57%)	6159 (46.46%)
Gabapentinoids	1823 (4.23%)	118 (6.47%)	129 (7.08%)	445 (24.41%)	485 (26.60%)	646 (35.44%)
Antidepressants	6883 (15.96%)	346 (5.03%)	325 (4.72%)	1555 (22.59%)	1748 (25.40%)	2909 (42.26%)
Polypharmacy[d]						
0	9404 (21.80%)	205 (2.18%)	246 (2.62%)	1294 (13.76%)	2418 (25.71%)	5241 (55.73%)
1-4	21008 (48.71%)	514 (2.45%)	548 (2.61%)	3275 (15.59%)	5548 (26.41%)	11123 (52.95%)
5-9	8880 (20.59%)	331 (3.73%)	339 (3.82%)	1722 (19.39%)	2283 (25.71%)	4205 (47.35%)

≥10	3838 (8.90%)	211 (5.50%)	200 (5.21%)	981 (25.56%)	905 (23.58%)	1541 (40.15%)
Smoking Status						
Never	16899 (39.18%)	393 (2.33%)	472 (2.79%)	2775 (16.42%)	4368 (25.85%)	8891 (52.61%)
Current	13961 (32.37%)	531 (3.80%)	498 (3.57%)	2667 (19.10%)	3523 (25.23%)	6742 (48.29%)
Ex Smoker	6095 (14.13%)	194 (3.18%)	205 (3.36%)	971 (15.93%)	1681 (27.58%)	3044 (49.94%)
Not Recorded	6175 (14.32%)	143 (2.32%)	158 (2.56%)	859 (13.91%)	1582 (25.62%)	3433 (55.60%)
BMI[e]						
Underweight/Normal	9010 (20.89%)	239 (2.65%)	269 (2.99%)	1474 (16.36%)	2275 (25.25%)	4753 (52.75%)
Overweight	9597 (22.25%)	282 (2.94%)	293 (3.05%)	1670 (17.40%)	2599 (27.08%)	4753 (49.53%)
Obese	9093 (21.08%)	328 (3.61%)	327 (3.60%)	1777 (19.54%)	2321 (25.53%)	4340 (47.73%)
Not Recorded	15430 (35.78%)	412 (2.67%)	444 (2.88%)	2351 (15.24%)	3959 (25.66%)	8264 (53.56%)
Modified CCI Score[f]						
0	35839 (83.10%)	928 (2.59%)	1061 (2.96%)	5800 (16.18%)	9353 (26.10%)	18697 (52.17%)
1	5763 (13.36%)	254 (4.41%)	207 (3.59%)	1138 (19.75%)	1425 (24.73%)	2739 (47.53%)
≥2	1528 (3.54%)	79 (5.17%)	65 (4.25%)	334 (21.86%)	376 (24.61%)	674 (44.11%)
Contact Count[g]						
1-10	12028 (27.89%)	227 (1.89%)	296 (2.46%)	1539 (12.80%)	3046 (25.32%)	6920 (57.53%)
11-15	6951 (16.12%)	175 (2.52%)	175 (2.52%)	1037 (14.92%)	1878 (27.02%)	3686 (53.03%)
16-25	10153 (23.54%)	283 (2.79%)	273 (2.69%)	1742 (17.16%)	2680 (26.40%)	5175 (50.97%)
≥26	13998 (32.46%)	576 (4.11%)	589 (4.21%)	2954 (21.10%)	3550 (25.36%)	6329 (45.21%)

Values are presented as n (column %) for the Overall column, and n (row %) by trajectory classes

Abbreviations: IMD = Index of Multiple Deprivation; MSK = Musculoskeletal; NSAIDs = Non-Steroidal Anti-Inflammatory Drugs; BMI = Body Mass Index; CCI = Charlson Comorbidity Index

[a] North of England defined as: Northeast, Northwest, Yorkshire and the Humber; Middle of England: East Midlands, West Midlands, East of England; and South of England: Southeast, Southwest, London

[b] Quintiles are used for IMD (1-5), where a higher score represents more deprived areas

[c] Excluding the index MSK consultation

[d] Excluding Opioids, NSAIDs, Gabapentinoids, Antidepressants
[e] Underweight/Normal: 10<=BMI<25; Overweight: 25<=BMI<30; Obese: 30<=BMI<80; Not Recorded: no BMI data available or BMI<10 or BMI>=80
[f] Excluding Rheumatic Disease
[g] Count of all medical consultations (except MSK related), as well as any recording of data such as BMI, smoking etc.

8.4.2 Multinomial Logistic Regression Results

There was presence of some correlation between the contact count and polypharmacy characteristics. Pearson's Correlation Coefficient was 0.71, and further, 91.6% of the people in the excessive polypharmacy category (≥10 drugs prescribed) were in the highest contact count category (≥26). Therefore, contact count was not used in the multivariable analyses, only polypharmacy.

The adjusted OR results of association between characteristics and the derived optimal trajectories for the index MSK condition fit note cohort are shown in Table 41.

Statistically significant associations were observed for both of the most severe work absence trajectories ('Chronic Sustained' and 'Chronic Fast Decreasing'), as compared to the reference 'Single' trajectory, with individuals in the more severe trajectories more likely to:

- Be older (either in the 46-55 or 56-66 years categories), as compared to a reference group of 16-25 year olds
- Live in the North of England or the Midlands, compared to the South
- Live in the most deprived areas of England (IMD = 5), compared to the least deprived areas (IMD = 1)
- Have more (≥2) MSK consultations in the two years prior to index fit note, compared to 0 MSK consultations
- Have baseline knee pain, osteoarthritis, or hip pain, compared to back pain
- Be prescribed opioids, gabapentinoids, or antidepressants in the two years prior to index fit note, compared to not being prescribed any of these treatments
- Be current smokers, compared to 'never' smokers

These results generally mirrored those of the univariable analyses, except being prescribed NSAIDs or being obese were no longer statistically significant.

Next, when comparing the two most severe work absence classes, there were a few instances of differences in significant observed associations, albeit such differences often occurred with point estimates that were close to the point of non-significance (i.e. the 95% CI of the OR was close to 1). Some examples of these differences included individuals:

- That were female, for the 'Chronic Fast Decreasing' class only
- That were young-middle aged (26-35 or 36-45 years), for the 'Chronic Sustained' class only. There was an increasing effect of association with higher age for all four non-reference age categories (26-35, 36-45, 46-55, 56-66 years).
- Living in the second-to-worst most deprived areas of England (IMD = 4), for the 'Chronic Sustained' class only
- With one MSK consultation in the two years prior to index fit note, for the 'Chronic Sustained' class only
- With excessive polypharmacy (≥10 drugs prescribed in the two years prior to index fit note, excluding opioids, NSAIDs, gabapentinoids, or antidepressants), compared to 0 drugs prescribed, for the 'Chronic Sustained' class only
- That were ex-smokers, for the 'Chronic Sustained' class only
- With a modified CCI score of 1 compared to a score of 0, for the 'Chronic Sustained' class only

The statistically significant associations with the highest magnitude in this analysis occurred for:

- The 56-66 year old age group, who had an association with OR (95% CI) of 3.26 (2.47, 4.30) of belonging to the ‘Chronic Sustained’ compared to ‘Single’ class, and 2.38 (1.83, 3.10) of belonging to the ‘Chronic Fast Decreasing’ compared to ‘Single’ class

As for the more severe absence trajectories, those in the ‘Intermittent Low’ group were more likely than those in the ‘Single’ trajectory to: be female (as for the ‘Chronic Fast Decreasing’ class only), live in the Midlands or more deprived areas of England, have more prior MSK consultations, have baseline knee pain, osteoarthritis, or hip pain, be prescribed opioids, gabapentinoids, or antidepressants, have excessive polypharmacy (as for the ‘Chronic Sustained’ class only), and to be current smokers.

But by contrast to the more severe absence trajectories, those in the ‘Intermittent Low’ class were also more likely to: not be aged 36-55 years (a protective effect was observed for these age groups), be prescribed NSAIDs, and to be overweight or obese.

Table 41. Characteristics Associated with Optimal Trajectories of Work Absence Due to a MSK Condition Using the 'Single' Trajectory Class as the Reference (Adjusted Model)

	Trajectory Class			
	Chronic Sustained[a] n=1,261	**Chronic Fast Decreasing[a]** n=1,333	**Intermittent Low[a]** n=7,272	**Short Term[a]** n=11,154
Sex				
Male	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)
Female	0.93 (0.81, 1.07)	**1.17 (1.01, 1.37)**	**1.15 (1.06, 1.24)**	**1.10 (1.03, 1.17)**
Age				
16-25 years	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)
26-35 years	**1.38 (1.02, 1.87)**	1.04 (0.79, 1.37)	1.00 (0.88, 1.14)	**1.24 (1.13, 1.36)**
36-45 years	**1.67 (1.25, 2.23)**	1.15 (0.88, 1.50)	**0.84 (0.74, 0.95)**	**1.15 (1.05, 1.26)**
46-55 years	**2.10 (1.59, 2.77)**	**1.42 (1.10, 1.85)**	**0.83 (0.73, 0.93)**	**1.37 (1.24, 1.50)**
56-66 years	**3.26 (2.47, 4.30)**	**2.38 (1.83, 3.10)**	1.04 (0.90, 1.20)	**1.54 (1.38, 1.71)**
Region[b]				
South of England	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)
North of England	**1.36 (1.16, 1.60)**	**1.70 (1.45, 2.00)**	1.02 (0.92, 1.13)	**1.36 (1.27, 1.46)**
Middle of England	**1.45 (1.24, 1.70)**	**1.24 (1.02, 1.51)**	**1.13 (1.03, 1.25)**	**1.20 (1.11, 1.29)**
IMD[c]				
1	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)
2	1.02 (0.78, 1.34)	1.17 (0.90, 1.54)	1.15 (0.99, 1.34)	1.02 (0.92, 1.12)
3	1.22 (0.94, 1.60)	1.19 (0.91, 1.55)	**1.32 (1.15, 1.52)**	1.01 (0.92, 1.12)
4	**1.68 (1.30, 2.16)**	1.27 (0.97, 1.66)	**1.61 (1.41, 1.84)**	0.99 (0.90, 1.09)

5	**2.50 (1.97, 3.17)**	**1.64 (1.28, 2.09)**	**1.78 (1.55, 2.04)**	0.95 (0.87, 1.05)
Missing	1.22 (0.79, 1.90)	**0.38 (0.15, 0.97)**	0.75 (0.53, 1.07)	**1.30 (1.12, 1.52)**
MSK Consultations - Prior 2 Years[d]				
0	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)
1	**1.27 (1.07, 1.51)**	1.06 (0.87, 1.29)	1.06 (0.96, 1.17)	**1.09 (1.01, 1.17)**
2	**1.63 (1.30, 2.03)**	**1.70 (1.35, 2.14)**	**1.29 (1.13, 1.47)**	**1.21 (1.09, 1.35)**
≥3	**1.81 (1.44, 2.27)**	**1.48 (1.16, 1.90)**	**1.47 (1.29, 1.67)**	**1.3 (1.17, 1.44)**
Baseline MSK Condition				
Back pain	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)
Knee pain	**1.34 (1.11, 1.62)**	**1.47 (1.21, 1.79)**	**1.34 (1.21, 1.49)**	0.95 (0.87, 1.03)
Hand/wrist pain	1.01 (0.72, 1.41)	1.21 (0.87, 1.69)	1.02 (0.86, 1.22)	**0.81 (0.71, 0.93)**
Inflammatory MSK	1.09 (0.80, 1.49)	0.95 (0.66, 1.38)	1.06 (0.86, 1.30)	**0.78 (0.66, 0.92)**
Osteoarthritis	**2.60 (2.00, 3.39)**	**1.84 (1.33, 2.53)**	**2.10 (1.70, 2.59)**	0.87 (0.72, 1.05)
Hip pain	**2.62 (2.00, 3.43)**	**1.68 (1.18, 2.39)**	**1.43 (1.15, 1.78)**	1.13 (0.95, 1.34)
Opioids				
No	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)
Yes	**1.40 (1.20, 1.65)**	**1.58 (1.32, 1.90)**	**1.18 (1.08, 1.30)**	**1.14 (1.06, 1.23)**
NSAIDs				
No	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)
Yes	1.16 (1.00, 1.35)	1.08 (0.91, 1.28)	**1.15 (1.05, 1.26)**	1.06 (0.99, 1.14)
Gabapentinoids				
No	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)
Yes	**1.60 (1.22, 2.11)**	**2.08 (1.58, 2.75)**	**1.51 (1.24, 1.83)**	**1.25 (1.05, 1.49)**

Antidepressants				
No	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)
Yes	**1.79 (1.52, 2.10)**	**1.48 (1.23, 1.77)**	**1.50 (1.35, 1.67)**	1.07 (0.98, 1.17)
Polypharmacy[e]				
0	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)
1-4	0.97 (0.80, 1.17)	0.82 (0.67, 1.01)	1.07 (0.96, 1.20)	0.99 (0.92, 1.07)
5-9	1.26 (1.00, 1.59)	1.14 (0.89, 1.45)	**1.36 (1.19, 1.55)**	1.01 (0.91, 1.12)
≥10	**1.48 (1.13, 1.95)**	1.37 (1.00, 1.88)	**1.84 (1.54, 2.20)**	1.00 (0.86, 1.15)
Smoking Status				
Never	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)
Current	**1.80 (1.53, 2.10)**	**1.41 (1.19, 1.69)**	**1.36 (1.25, 1.49)**	1.07 (1.00, 1.15)
Ex Smoker	**1.24 (1.01, 1.52)**	1.17 (0.94, 1.46)	0.99 (0.87, 1.11)	1.08 (0.99, 1.17)
Not Recorded	1.22 (0.95, 1.57)	1.03 (0.79, 1.34)	0.95 (0.83, 1.08)	1.00 (0.91, 1.10)
BMI[f]				
Underweight/Normal	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)
Overweight	0.95 (0.78, 1.17)	0.94 (0.76, 1.18)	**1.16 (1.04, 1.30)**	**1.12 (1.02, 1.22)**
Obese	0.99 (0.82, 1.21)	0.97 (0.78, 1.20)	**1.19 (1.07, 1.33)**	1.05 (0.96, 1.14)
Not Recorded	1.20 (0.98, 1.48)	1.11 (0.90, 1.37)	1.07 (0.96, 1.20)	1.06 (0.98, 1.15)
Modified CCI Score[g]				
0	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)
1	**1.34 (1.12, 1.61)**	0.90 (0.72, 1.13)	1.01 (0.90, 1.14)	0.96 (0.88, 1.05)
≥2	1.28 (0.95, 1.72)	0.89 (0.62, 1.29)	1.10 (0.91, 1.34)	0.95 (0.80, 1.13)

Abbreviations: IMD = Index of Multiple Deprivation; MSK = Musculoskeletal; NSAIDs = Non-Steroidal Anti-Inflammatory Drugs; BMI = Body Mass Index; CCI = Charlson Comorbidity Index

Notes: Values are presented as adjusted odds ratios with 95% confidence intervals (adjustments were made for all the other variables shown in the Table)

Statistically significant estimates (where 95% CI doesn't include the value 1) are shown in bold

[a] All odds ratios are calculated with respect to the reference trajectory: Single.
[b] North of England defined as: Northeast, Northwest, Yorkshire and the Humber; Middle of England: East Midlands, West Midlands, East of England; and South of England: Southeast, Southwest, London
[c] Quintiles are used for IMD (1-5), where a higher score represents more deprived areas
[d] Excluding the index MSK consultation
[e] Excluding Opioids, NSAIDs, Gabapentinoids, Antidepressants
[f] Underweight/Normal: 10<=BMI<25; Overweight: 25<=BMI<30; Obese: 30<=BMI<80; Not Recorded: no BMI data available or BMI<10 or BMI>=80
[g] Excluding Rheumatic Disease

8.5 Results: Characteristics Associated with Optimal Trajectories of Work Absence Due to a MH Condition

8.5.1 Descriptive Summary of Optimal Trajectories and Univariable Associations

As in the MSK condition analysis, the profiles of the two most severe work absence trajectories ('Chronic Sustained' and 'Chronic Fast Decreasing') for the index MH condition fit note cohort were also highly similar (Table 42).

In particular, the profiles of these two trajectory classes consisted of relatively greater numbers of individuals:

- That were male
- That were older (46-66 years)
- Living in the North of England
- Living in the most deprived areas of England (IMD = 5)
- With more (≥1) MH consultations in the two years prior to index fit note
- With more (≥3) MH consultations during the three-year follow-up
- With baseline depression, or anxiety and depression combined
- Prescribed opioids, NSAIDs, gabapentinoids, or antidepressants in the two years prior to index fit note
- With excessive polypharmacy (≥10 drugs prescribed in the two years prior to index fit note, excluding opioids, NSAIDs, gabapentinoids, or antidepressants)
- That were current smokers
- That were obese or had a 'not recorded' BMI category
- With at least one other comorbidity (in addition to a MH condition) in the two years prior to index fit note date (CCI score of ≥1)

- With a greater contact count (≥26) in the two years prior to index fit note date

The few differences between these two class profiles now were that the ‘Chronic Sustained’ class also consisted of relatively more individuals aged 16-25 years, living in the Midlands or the second-to-worst most deprived areas of England (IMD=4), and with 5-9 drugs prescribed in the two years prior to index fit note date (the ‘Chronic Fast Decreasing’ class did not).

Whilst the ‘Chronic Fast Decreasing’ class also had relatively more individuals with 0 drugs prescribed in the two years prior to index fit note date (the ‘Chronic Sustained’ class did not).

The profile of the ‘Intermittent Low’ class was again most similar to these two most severe work absence classes, although there were a few differences, such as that the ‘Intermittent Low’ class contained relatively more females, and individuals living in the South of England.

In contrast, the profile for the ‘Single’ class remained largely opposite to the three classes described thus far and was similar to the ‘Short Term’ class for this index MH condition fit note cohort.

The univariable associations of characteristics are shown in Appendix Table N.2. These again mainly confirmed the descriptive profiles of the two most severe absence trajectories.

Table 42. Descriptive Statistics of Optimal Trajectories of Work Absence Due to a MH Condition

	Overall	Trajectory Class				
		Chronic Sustained	Chronic Fast Decreasing	Intermittent Low	Short Term	Single
	n=62,355	2881 (4.62%)	3848 (6.17%)	10835 (17.38%)	21534 (34.53%)	23257 (37.30%)
Female	36937 (59.24%)	1560 (4.22%)	2143 (5.80%)	6625 (17.94%)	12618 (34.16%)	13991 (37.88%)
Age						
16-25 years	17271 (27.70%)	845 (4.89%)	1000 (5.79%)	3546 (20.53%)	4997 (28.93%)	6883 (39.85%)
26-35 years	16700 (26.78%)	615 (3.68%)	866 (5.19%)	2856 (17.10%)	5895 (35.3%)	6468 (38.73%)
36-45 years	12952 (20.77%)	555 (4.29%)	792 (6.11%)	1966 (15.18%)	4880 (37.68%)	4759 (36.74%)
46-55 years	10562 (16.94%)	545 (5.16%)	781 (7.39%)	1637 (15.50%)	3997 (37.84%)	3602 (34.10%)
56-66 years	4870 (7.81%)	321 (6.59%)	409 (8.40%)	830 (17.04%)	1765 (36.24%)	1545 (31.72%)
Region[a]						
North of England	19310 (30.97%)	972 (5.03%)	1445 (7.48%)	3337 (17.28%)	6883 (35.64%)	6673 (34.56%)
Middle of England	14487 (23.23%)	782 (5.40%)	911 (6.29%)	2416 (16.68%)	5103 (35.22%)	5275 (36.41%)
South of England	28558 (45.80%)	1127 (3.95%)	1492 (5.22%)	5082 (17.80%)	9548 (33.43%)	11309 (39.60%)
IMD[b]						
1	9903 (15.88%)	291 (2.94%)	447 (4.51%)	1391 (14.05%)	3785 (38.22%)	3989 (40.28%)
2	10623 (17.04%)	376 (3.54%)	564 (5.31%)	1631 (15.35%)	3861 (36.35%)	4191 (39.45%)
3	10834 (17.37%)	458 (4.23%)	638 (5.89%)	1837 (16.96%)	3757 (34.68%)	4144 (38.25%)
4	12514 (20.07%)	623 (4.98%)	789 (6.30%)	2393 (19.12%)	4143 (33.11%)	4566 (36.49%)
5	15627 (25.06%)	1071 (6.85%)	1290 (8.25%)	3287 (21.03%)	4794 (30.68%)	5185 (33.18%)
Missing	2854 (4.58%)	62 (2.17%)	120 (4.20%)	296 (10.37%)	1194 (41.84%)	1182 (41.42%)
MH Consultations - Prior 2 Years[c]						

0	36827 (59.06%)	1514 (4.11%)	2143 (5.82%)	5722 (15.54%)	13385 (36.35%)	14063 (38.19%)
1	10925 (17.52%)	527 (4.82%)	711 (6.51%)	1999 (18.30%)	3643 (33.35%)	4045 (37.03%)
2	5225 (8.38%)	291 (5.57%)	364 (6.97%)	1015 (19.43%)	1681 (32.17%)	1874 (35.87%)
≥3	9378 (15.04%)	549 (5.85%)	630 (6.72%)	2099 (22.38%)	2825 (30.12%)	3275 (34.92%)
MH Consultations - 3 Year Follow-up[c]						
0	16689 (26.76%)	205 (1.23%)	389 (2.33%)	1727 (10.35%)	3511 (21.04%)	10857 (65.05%)
1	10995 (17.63%)	185 (1.68%)	274 (2.49%)	1539 (14.00%)	4644 (42.24%)	4353 (39.59%)
2	7708 (12.36%)	156 (2.02%)	280 (3.63%)	1385 (17.97%)	3519 (45.65%)	2368 (30.72%)
≥3	26963 (43.24%)	2335 (8.66%)	2905 (10.77%)	6184 (22.94%)	9860 (36.57%)	5679 (21.06%)
Baseline MH Condition						
Stress	16339 (26.20%)	395 (2.42%)	656 (4.01%)	2348 (14.37%)	5851 (35.81%)	7089 (43.39%)
Anxiety and Depression	14769 (23.69%)	896 (6.07%)	1198 (8.11%)	2853 (19.32%)	5136 (34.78%)	4686 (31.73%)
Depression	15695 (25.17%)	952 (6.07%)	1189 (7.58%)	2984 (19.01%)	5381 (34.28%)	5189 (33.06%)
Anxiety	15552 (24.94%)	638 (4.10%)	805 (5.18%)	2650 (17.04%)	5166 (33.22%)	6293 (40.46%)
Opioids	8139 (13.05%)	545 (6.70%)	618 (7.59%)	1844 (22.66%)	2484 (30.52%)	2648 (32.53%)
NSAIDs	8023 (12.87%)	429 (5.35%)	514 (6.41%)	1577 (19.66%)	2674 (33.33%)	2829 (35.26%)
Gabapentinoids	1122 (1.80%)	82 (7.31%)	87 (7.75%)	284 (25.31%)	309 (27.54%)	360 (32.09%)
Antidepressants	24934 (39.99%)	1422 (5.70%)	1726 (6.92%)	5202 (20.86%)	7806 (31.31%)	8778 (35.20%)
Polypharmacy[d]						
0	12927 (20.73%)	604 (4.67%)	846 (6.54%)	1959 (15.15%)	4657 (36.03%)	4861 (37.60%)
1-4	31262 (50.14%)	1309 (4.19%)	1804 (5.77%)	5091 (16.28%)	11082 (35.45%)	11976 (38.31%)
5-9	13375 (21.45%)	648 (4.84%)	821 (6.14%)	2591 (19.37%)	4432 (33.14%)	4883 (36.51%)
≥10	4791 (7.68%)	320 (6.68%)	377 (7.87%)	1194 (24.92%)	1363 (28.45%)	1537 (32.08%)
Smoking Status						
Never	23727 (38.05%)	859 (3.62%)	1219 (5.14%)	3804 (16.03%)	8351 (35.20%)	9494 (40.01%)

Current	21566 (34.59%)	1335 (6.19%)	1656 (7.68%)	4284 (19.86%)	6974 (32.34%)	7317 (33.93%)
Ex Smoker	8171 (13.10%)	313 (3.83%)	452 (5.53%)	1304 (15.96%)	3034 (37.13%)	3068 (37.55%)
Not Recorded	8891 (14.26%)	374 (4.21%)	521 (5.86%)	1443 (16.23%)	3175 (35.71%)	3378 (37.99%)
BMI[e]						
Underweight/Normal	17637 (28.28%)	724 (4.11%)	987 (5.60%)	3093 (17.54%)	5972 (33.86%)	6861 (38.90%)
Overweight	11454 (18.37%)	488 (4.26%)	656 (5.73%)	1886 (16.47%)	4128 (36.04%)	4296 (37.51%)
Obese	10218 (16.39%)	554 (5.42%)	697 (6.82%)	1959 (19.17%)	3548 (34.72%)	3460 (33.86%)
Not Recorded	23046 (36.96%)	1115 (4.84%)	1508 (6.54%)	3897 (16.91%)	7886 (34.22%)	8640 (37.49%)
CCI Score						
0	52355 (83.96%)	2318 (4.43%)	3195 (6.10%)	8739 (16.69%)	18306 (34.97%)	19797 (37.81%)
1	8323 (13.35%)	462 (5.55%)	531 (6.38%)	1728 (20.76%)	2687 (32.28%)	2915 (35.02%)
≥2	1677 (2.69%)	101 (6.02%)	122 (7.27%)	368 (21.94%)	541 (32.26%)	545 (32.50%)
Contact Count[f]						
1-10	15410 (24.71%)	619 (4.12%)	912 (6.07%)	2126 (14.14%)	5481 (36.45%)	5899 (39.23%)
11-15	10003 (16.04%)	393 (4.05%)	603 (6.21%)	1501 (15.45%)	3547 (36.51%)	3670 (37.78%)
16-25	15061 (24.15%)	664 (4.54%)	882 (6.03%)	2599 (17.77%)	4970 (33.97%)	5514 (37.69%)
≥26	21881 (35.09%)	1155 (5.45%)	1441 (6.80%)	4544 (21.45%)	6683 (31.54%)	7364 (34.76%)

Values are presented as n (column %) for the Overall column, and n (row %) by trajectory classes

Abbreviations: IMD = Index of Multiple Deprivation; MH = Mental Health; NSAIDs = Non-Steroidal Anti-Inflammatory Drugs; BMI = Body Mass Index; CCI = Charlson Comorbidity Index

[a] North of England defined as: Northeast, Northwest, Yorkshire and the Humber; Middle of England: East Midlands, West Midlands, East of England; and South of England: Southeast, Southwest, London

[b] Quintiles are used for IMD (1-5), where a higher score represents more deprived areas

[c] Excluding the index MH consultation

[d] Excluding Opioids, NSAIDs, Gabapentinoids, Antidepressants

[e] Underweight/Normal: 10<=BMI<25; Overweight: 25<=BMI<30; Obese: 30<=BMI<80; Not Recorded: no BMI data available or BMI<10 or BMI>=80

[f] Count of all medical consultations (except MH related), as well as any recording of data such as BMI, smoking etc.

8.5.2 Multinomial Logistic Regression Results

As with the index MSK condition fit note cohort, presence of correlation between the contact count and polypharmacy characteristics was also apparent for the index MH condition fit note cohort. Pearson's Correlation Coefficient was similar at 0.68, and 92.7% of those with excessive polypharmacy (≥10 drugs prescribed) category data were in the highest contact count category (≥26). Thus, only polypharmacy was retained in the multivariable analyses again, with full results shown in Table 43.

The statistically significant associations for the two most severe work absence trajectories ('Chronic Sustained' and 'Chronic Fast Decreasing'), compared to the reference 'Single' trajectory, were that individuals in the more severe trajectories were more likely to:

- Be male
- Be older (in the: 36-45, 46-55, or 56-66 years categories), as compared to a reference group of 16-25 year olds. An increasing association with higher age was shown.
- Live in the North of England or the Midlands, compared to the South
- Live in any of the three most deprived areas of England (IMD from 3 to 5), compared to the least deprived areas (IMD = 1). An increasing association with higher IMD values was shown.
- Have baseline anxiety, depression, or anxiety and depression combined, compared to stress, with the strongest association with anxiety and depression combined, followed by depression alone.
- Be prescribed an opioid in the prior two years

- Have excessive polypharmacy (≥10 drugs prescribed in the two years prior to index fit note, excluding opioids, NSAIDs, gabapentinoids, or antidepressants), compared to 0 drugs prescribed
- Be current smokers, compared to 'never' smokers
- Be obese or with a 'not recorded' BMI, compared to 'underweight/normal' BMI

As with the MSK condition trajectory-covariate association analyses, these multivariable association results for the MH condition fit note cohort also largely mirrored the univariable analyses results.

There were again few differences in significantly associated characteristics between the two most severe work absence classes in the multivariable analysis. Two of these differences were associations that were significant for the 'Chronic Sustained' class only, and concerned individuals that were: 26-35 years old (a protective effect), or with ≥3 MH consultations in the two years prior to the index fit note.

The only association that was significant for the 'Chronic Fast Decreasing' class and not the 'Chronic Sustained', was with individuals living in a neighbourhood with an IMD value of 2 (the second best).

The strongest statistically significant associations observed overall, were for:

- Individuals with baseline anxiety and depression combined, as opposed to stress, with an OR (95% CI) of 3.72 (3.17, 4.36) of belonging to the 'Chronic Sustained' compared to 'Single' class, and 3.31 (2.89, 3.80) of belonging to the 'Chronic Fast Decreasing' compared to 'Single' class

- The association of baseline depression, compared to stress, was next highest (3.32 (2.85, 3.87) and 2.77 (2.42, 3.18), for the ‘Chronic Sustained’ and ‘Chronic Fast Decreasing’ classes, respectively)

As for the more severe absence trajectories, those in the ‘Intermittent Low’ group were more likely than those in the ‘Single’ trajectory to: not be aged 26-35 years (a protective effect was observed for this age group, as for the ‘Chronic Sustained’ class only), live in more deprived areas of England, have more prior MH consultations (as for the ‘Chronic Sustained’ class only), have baseline anxiety, depression, or anxiety and depression combined, be prescribed opioids, have excessive polypharmacy, be a current smoker and to be obese.

But by contrast to the more severe absence trajectories, those in the ‘Intermittent Low’ class were also more likely to: not be aged 36-45 years (a protective effect was observed for this age group), be prescribed antidepressants, and to have at least one other comorbidity in addition to a MH condition (CCI score ≥1).

Table 43. Characteristics Associated with Optimal Trajectories of Work Absence Due to a MH Condition Using the 'Single' Trajectory Class as the Reference

	Trajectory Class			
	Chronic Sustained[a]	**Chronic Fast Decreasing[a]**	**Intermittent Low[a]**	**Short Term[a]**
	n=2,881	n=3,848	n=10,835	n=21,534
Sex				
Male	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)
Female	**0.79 (0.71, 0.87)**	**0.88 (0.80, 0.96)**	1.00 (0.93, 1.07)	0.98 (0.93, 1.03)
Age				
16-25 years	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)
26-35 years	**0.83 (0.73, 0.95)**	1.05 (0.92, 1.19)	**0.85 (0.78, 0.93)**	**1.38 (1.30, 1.47)**
36-45 years	**1.18 (1.04, 1.35)**	**1.52 (1.33, 1.74)**	**0.79 (0.72, 0.88)**	**1.63 (1.52, 1.74)**
46-55 years	**1.66 (1.43, 1.92)**	**2.17 (1.90, 2.49)**	0.91 (0.82, 1.01)	**1.80 (1.66, 1.94)**
56-66 years	**2.30 (1.92, 2.77)**	**2.72 (2.29, 3.23)**	1.08 (0.94, 1.24)	**1.87 (1.69, 2.08)**
Region[b]				
South of England	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)
North of England	**1.31 (1.15, 1.49)**	**1.65 (1.48, 1.83)**	0.99 (0.91, 1.08)	**1.34 (1.26, 1.42)**
Middle of England	**1.51 (1.33, 1.72)**	**1.34 (1.20, 1.50)**	0.95 (0.88, 1.03)	**1.20 (1.13, 1.27)**
IMD[c]				
1	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)
2	1.21 (1.00, 1.47)	**1.19 (1.01, 1.41)**	1.12 (1.00, 1.26)	0.96 (0.89, 1.04)
3	**1.49 (1.24, 1.79)**	**1.36 (1.16, 1.60)**	**1.28 (1.14, 1.43)**	0.95 (0.88, 1.02)
4	**1.86 (1.55, 2.23)**	**1.49 (1.27, 1.76)**	**1.52 (1.36, 1.71)**	0.95 (0.88, 1.03)
5	**2.67 (2.24, 3.18)**	**2.00 (1.71, 2.33)**	**1.83 (1.64, 2.04)**	0.93 (0.86, 1.01)

Missing	**0.68 (0.48, 0.95)**	0.93 (0.72, 1.19)	**0.54 (0.41, 0.71)**	1.11 (0.99, 1.25)
MH Consultations - Prior 2 Years[d]				
0	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)
1	1.03 (0.90, 1.17)	1.07 (0.94, 1.22)	1.08 (0.98, 1.18)	0.97 (0.91, 1.04)
2	1.17 (0.98, 1.41)	1.14 (0.95, 1.36)	1.12 (0.99, 1.26)	0.99 (0.90, 1.09)
≥3	**1.20 (1.02, 1.42)**	1.06 (0.91, 1.25)	**1.28 (1.15, 1.43)**	0.94 (0.86, 1.03)
Baseline MH Condition				
Stress	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)
Anxiety and Depression	**3.72 (3.17, 4.36)**	**3.31 (2.89, 3.80)**	**1.69 (1.53, 1.86)**	**1.63 (1.52, 1.75)**
Depression	**3.32 (2.85, 3.87)**	**2.77 (2.42, 3.18)**	**1.55 (1.40, 1.70)**	**1.50 (1.40, 1.60)**
Anxiety	**2.04 (1.74, 2.38)**	**1.56 (1.35, 1.80)**	**1.19 (1.09, 1.31)**	**1.13 (1.06, 1.21)**
Opioids				
No	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)
Yes	**1.37 (1.20, 1.56)**	**1.16 (1.01, 1.33)**	**1.33 (1.21, 1.46)**	0.97 (0.89, 1.04)
NSAIDs				
No	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)
Yes	1.06 (0.92, 1.22)	0.97 (0.85, 1.11)	1.04 (0.95, 1.14)	1.02 (0.95, 1.10)
Gabapentinoids				
No	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)
Yes	0.98 (0.73, 1.32)	0.87 (0.62, 1.22)	1.14 (0.91, 1.43)	0.82 (0.67, 1.01)
Antidepressants				
No	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)
Yes	1.09 (0.97, 1.23)	0.95 (0.85, 1.07)	**1.25 (1.14, 1.36)**	**0.85 (0.80, 0.90)**
Polypharmacy[e]				
0	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)

1-4	0.93 (0.82, 1.06)	0.93 (0.83, 1.04)	1.05 (0.95, 1.15)	1.00 (0.94, 1.07)
5-9	1.03 (0.88, 1.21)	1.01 (0.87, 1.17)	**1.22 (1.09, 1.36)**	0.99 (0.92, 1.07)
≥10	**1.27 (1.02, 1.59)**	**1.29 (1.05, 1.59)**	**1.59 (1.36, 1.86)**	0.96 (0.85, 1.09)
Smoking Status				
Never	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)
Current	**1.75 (1.56, 1.95)**	**1.53 (1.38, 1.70)**	**1.44 (1.33, 1.55)**	1.05 (0.99, 1.11)
Ex Smoker	0.99 (0.84, 1.17)	0.99 (0.85, 1.14)	1.06 (0.96, 1.18)	1.04 (0.96, 1.11)
Not Recorded	1.17 (1.00, 1.38)	1.11 (0.96, 1.28)	**1.2 (1.08, 1.33)**	**1.09 (1.01, 1.17)**
BMI[f]				
Underweight/Normal	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)
Overweight	1.00 (0.87, 1.16)	0.97 (0.85, 1.11)	0.97 (0.88, 1.07)	1.04 (0.97, 1.11)
Obese	**1.28 (1.11, 1.48)**	**1.21 (1.05, 1.39)**	**1.19 (1.08, 1.31)**	**1.12 (1.04, 1.21)**
Not Recorded	**1.22 (1.07, 1.39)**	**1.18 (1.04, 1.33)**	1.06 (0.97, 1.15)	1 (0.94, 1.06)
CCI Score				
0	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)
1	1.14 (0.99, 1.31)	0.96 (0.83, 1.10)	**1.21 (1.09, 1.33)**	0.97 (0.90, 1.05)
≥2	1.10 (0.84, 1.43)	0.99 (0.76, 1.29)	**1.34 (1.10, 1.63)**	0.98 (0.84, 1.15)

Abbreviations: IMD = Index of Multiple Deprivation; MH = Mental Health; NSAIDs = Non-Steroidal Anti-Inflammatory Drugs; BMI = Body Mass Index; CCI = Charlson Comorbidity Index

Notes: Values are presented as adjusted odds ratios with 95% confidence intervals (adjustments were made for all the other variables in the Table).
Statistically significant estimates (where 95% CI doesn't include the value 1) are shown in bold

[a] All odds ratios are calculated with respect to the reference trajectory: Single.

[b] North of England defined as: Northeast, Northwest, Yorkshire and the Humber; Middle of England: East Midlands, West Midlands, East of England; and South of England: Southeast, Southwest, London

[c] Quintiles are used for IMD (1-5), where a higher score represents more deprived areas

[d] Excluding the index MH consultation

[e] Excluding Opioids, NSAIDs, Gabapentinoids, Antidepressants

[f] Underweight/Normal: 10<=BMI<25; Overweight: 25<=BMI<30; Obese: 30<=BMI<80; Not Recorded: no BMI data available or BMI<10 or BMI>=80

8.6 Results: Sensitivity Analyses

In this Section, the trajectory-covariate association analyses were re-run under the same two sensitivity analyses from the previous Chapter, using other approaches to handling individuals with incomplete follow-up data:

- Sensitivity analyses 1: individuals were excluded if they met an exclusion reason at any point during the calendar year of the index fit note, without any censoring after index year.
- Sensitivity analyses 2: individuals were excluded if they met an exclusion reason at any point during one year follow-up post index fit note issuance

The results from these two sensitivity analyses, and the main analyses whereby individuals with incomplete follow-up data were retained, were highly similar, both in terms of the characteristics that emerged as statistically significantly associated with the more severe absence trajectories, as well as the magnitude and 95% CIs of these trajectory-characteristic associations.

As an example, for the MSK condition fit note cohort analysis, individuals living in the most deprived areas of England (IMD = 5) had an OR (95% CI) of 2.50 (1.97, 3.17) of belonging to the Chronic Sustained class in the main analysis. In sensitivity analysis 1, the OR (95% CI) was 2.49 (1.96, 3.16), and in sensitivity analysis 2, 2.50 (1.96, 3.17).

8.7 Discussion

8.7.1 Summary of Findings and Comparison with Other Research

In this study, a total of 16 characteristics were explored for association with the five optimal trajectories of work absence identified in the previous Chapter. For each of these five trajectory classes, descriptive profiles were first uncovered, based on

sociodemographic and health characteristics, types of treatment received, and comorbidity.

The typical profiles of the two most severe work absence trajectories ('Chronic Sustained' and 'Chronic Fast Decreasing') were highly similar, and distinct from the least severe work absence trajectory ('Single'). The 'Intermittent Low' trajectory class was also found to have a similar profile to these 'Chronic Sustained' and 'Chronic Fast Decreasing' trajectory classes. In contrast, the 'Short Term' class had a profile that was harder to categorise, involving elements of both the 'Single' and more adverse trajectories for the index MSK condition fit note cohort, but most similar to the 'Single' class for the index MH condition fit note cohort.

Aside from the 'Short Term' class, the remaining four classes were largely similar between the index MSK condition and MH condition fit note cohorts.

Then, multivariable multinomial logistic regression analysis was carried out to test association of these characteristics with the most severe work absence trajectories of future persistent or recurrent work absence, as opposed to the 'Single' trajectory. The characteristics that showed statistically significant associations for both the 'Chronic Sustained' and 'Chronic Fast Decreasing' classes are summarised in Table 44.

In brief, the characteristics significantly associated with both of the two most severe trajectory classes, that were identified in the MSK and MH condition index fit note cohorts, were that individuals in the more severe trajectories were more likely to be:

- Older (46-66 years of age)
- Living in the North of England or the Midlands
- Living in the most deprived areas of England

- Current smokers
- Prescribed opioids in the two years prior to index fit note

Then, only for the index MSK condition fit note cohort, additional significant associations were that individuals in the more severe trajectories were more likely to:

- Have ≥2 MSK consultations in the two years prior to index fit note
- Have baseline knee pain, osteoarthritis, or hip pain
- Be prescribed gabapentinoids or antidepressants in the two years prior to index fit note

And significant associations for only the index MH condition fit note cohort were that individuals in the more severe trajectories were more likely to:

- Be male
- Have baseline anxiety, depression, or anxiety and depression
- Be obese, or have a ‘not recorded’ BMI status
- Have excessive polypharmacy

Table 44. Summary of Characteristics Significantly Associated with Both of the Two Most Severe Work Absence Trajectories

		Cohort			
		MSK Condition Index Fit Note		MH Condition Index Fit Note	
Type of Characteristic	**Characteristic**	**Statistically Significant Association?**	**Significant Category**	**Statistically Significant Association?**	**Significant Category**
Sociodemographic	Sex	×	×	✓	Males
	Age	✓	46-55, 55-66 years	✓	36-45, 46-55, 55-66 years
	Region[a]	✓	North or Middle of England	✓	North or Middle of England
	IMD[b]	✓	Most Deprived Areas (IMD = 5)	✓	Most Deprived Areas (IMD = 3 to 5)
Health	MSK Consultations - Prior 2 Years[c]	✓	≥2	N/A	N/A
	MH Consultations - Prior 2 Years[d]	N/A	N/A	×	×
	Baseline MSK Condition	✓	Knee pain, Hip pain, Osteoarthritis	N/A	N/A
	Baseline MH Condition	N/A	N/A	✓	Anxiety, Depression, Anxiety and Depression
	Smoking Status	✓	Current Smokers	✓	Current Smokers
	BMI[e]	×	×	✓	Obese or 'Not Recorded'
	Opioids	✓	Yes	✓	Yes

Types of Treatment Received	NSAIDs	×	×	×	×
	Gabapentinoids	✓	Yes	×	×
	Antidepressants	✓	Yes	×	×
Comorbidity	Polypharmacy[f]	×	×	✓	≥10
	CCI Score	N/A	N/A	×	×
	Modified CCI Score[g]	×	×	N/A	N/A

Notes: A tick indicates that a statistically significant association was found, whereas a cross signifies it was not.
Abbreviations: N/A = Not Applicable; IMD = Index of Multiple Deprivation; MSK = Musculoskeletal; MH = Mental Health; NSAIDs = Non-Steroidal Anti-Inflammatory Drugs; BMI = Body Mass Index; CCI = Charlson Comorbidity Index
[a] North of England defined as: North East, North West, Yorkshire and the Humber; Middle of England: East Midlands, West Midlands, East of England; and South of England: South East, South West, London
[b] Quintiles are used for IMD (1-5), where a higher score represents more deprived areas
[c] Excluding the index MSK consultation
[d] Excluding the index MH consultation
[e] Underweight/Normal: 10<=BMI<25; Overweight: 25<=BMI<30; Obese: 30<=BMI<80; Not Recorded: no BMI data available or BMI<10 or BMI>=80
[f] Excluding Opioids, NSAIDs, Gabapentinoids, Antidepressants
[g] Excluding Rheumatic Disease

Face validity of these findings was affirmed by a GP through a presentation AL delivered on 12th October 2023. OHID also found these results plausible after AL presented to them on 31st January 2024, as did a PPIE group (AL disseminated findings to them on 1st February 2024).

The finding that older individuals are associated with the longer-term absence trajectories has been observed in other studies: Farrants et al (2018),[72] Farrants et al (2019),[69] Spronken et al,[73] and Rysstad et al.[101]

8.7.2 Study Strengths and Limitations

As with study 2 of this thesis regarding trajectory derivation, one of the key strengths of this study was that it was the first to explore association of characteristics with trajectories of work absence for an English population. Furthermore, as the same large sample of CPRD Aurum data from study 2 was retained in this study, the results again are expected to be representative of the general English population.[128]

However, whilst best efforts were taken to minimise the level of missing data during the process of defining the characteristics, one key limitation was the substantial amount of missing data for the BMI characteristic (35.8% and 37.0%, for the index MSK condition fit note and index MH condition fit note cohorts, respectively).

The ‘not recorded’ category used to define missing BMI, was found to be significantly associated with membership in the two most adverse trajectories of work absence, for the index MH condition fit note cohort. These results should be taken with caution, as it is not clear what the profile of these individuals with missing BMI represents. Albeit the observed association was only weakly significant with a low magnitude and a 95% CI close to 1 (OR (95% CI) of 1.22 (1.07, 1.39) and 1.18 (1.04, 1.33), for the ‘Chronic

Sustained' and 'Chronic Fast Decreasing' trajectories, respectively). As mentioned, multiple imputation was not a viable option for BMI due to the MAR assumption being unlikely to hold.

Whilst it was a strength of this study that a broad range of characteristics were tested for association within the confines of data available from CPRD, one of the other main limitations was that work-related characteristic data was not available through CPRD linkage. For example, it would have been beneficial to consider type of work (unskilled, manual, skilled etc), level of job control, emotional demands, interpersonal relations, and self-efficacy to RTW.

Self-efficacy to RTW, is often a strong predictor in studies assessing RTW after an absence due to a MSK or MH condition.[211–214] Indeed, in the study by Rysstad et al,[101] negative RTW expectancy was the only covariate associated with all three of their most adverse work absence trajectories.

8.7.3 Conclusion

In this first ever study of association of characteristics with future persistent or recurrent work absence in an English population, a set of characteristics were found to be associated with both of our two trajectory classes comprised of greater fit note issuance. These characteristics, identified for cohorts of individuals with an index fit note due to either a MSK or MH condition, were individuals: who were older, living in the North of England or the Midlands or the most deprived areas of England, prescribed opioids, and who were current smokers. These findings have implications for prevention and management strategies for individuals experiencing an incident work absence due to either a MSK or MH condition.

Chapter 9. Discussion

In this Chapter, the key findings from this thesis are first discussed and set into context (Section 9.1). Then, the key thesis strengths and limitations are explored in Section 9.2, and further work is recommended in Section 9.3. Finally, this thesis closes with a summary of final conclusions (Section 9.4).

9.1 Discussion of Thesis Findings

9.1.1 Overview of Thesis Findings

First, in Chapter 3, a systematic review was conducted to identify the extent of published literature concerning trajectories of work absence due to a MSK or MH condition. Work absence trajectory research was found to be scarce, with only six studies ultimately retained in the review (published from 2016-2023), and none from the UK.

There was considerable heterogeneity when comparing these studies, hence a meta-analysis was not performed. Instead, through a narrative synthesis, it was shown that whilst the trajectories identified across the six included studies did vary in number and prevalence, there were similarities in the trajectory shapes of the faster and slower RTW trajectories. Another important finding was that the faster RTW trajectories tended to occur with higher prevalence, and the slower RTW trajectories with lower prevalence. No clear differences were observed when comparing the faster and slower RTW trajectories by reason for index fit note (i.e., through comparing the three absence trajectory studies where the baseline absence was due to a MSK condition, against three due to a MH condition). Furthermore, a key theme from the included studies was that the general quality of reporting of the trajectory analysis was in need of improvement.

Thus, this motivated the need for a study of trajectories of work absence in a UK setting, and with higher quality trajectory reporting.

Prior to performing such a study, it was important to explore trends in incidence rates of fit note issuance due to a MSK or MH condition over 2010-2021 (Chapter 6). A large primary care database based on an English population, CPRD Aurum, was used for all studies in this thesis. Unlike the trajectories of absence in the systematic review, differences were observed when comparing incidence rates by MSK and MH condition fit notes in CPRD data (potential reasons for these differences were discussed earlier in Section 6.4.1). MSK condition fit note incidence showed a decreasing trend from 2014 to 2021, whereas for MH conditions there was an increasing fit note incidence over this period. Furthermore, for MSK conditions there were no clear differences in fit note incidence rates by sociodemographic variables, whereas for MH conditions, younger females and persons living in the Northeast and Northwest of England had a higher incidence.

Using a subset of individuals from the incidence rates study, trajectories of work absence were derived in Chapter 7. These trajectories were based on cohorts with an incident fit note during 2016-2018, either due to a MSK or MH condition. A binary fit note definition was used over follow-up, of 1/0 for fit notes issued due to any reason within a given time interval.

In line with the systematic review, the derived trajectory models were generally similar when comparing between the MSK and MH condition fit note cohorts. The chosen optimal model for both MSK and MH condition fit note cohorts was a five-class LCGA model that was based on two-monthly recurring intervals over a one-year follow-up since index date. When comparing trajectory derivation results across different

approaches to specifying interval and follow-up lengths, trajectory models based on a shorter-term follow-up of one-year, outperformed those based on a longer-term follow-up of three-years.

In the optimal trajectory models, two subgroups were identified that consisted of longer-term absence patterns ('Chronic Sustained' and 'Chronic Fast Decreasing' trajectory classes), and these occurred with low prevalence, in agreement with the systematic review findings. Furthermore, two subgroups characterised by less severe absence patterns were also found ('Single' and 'Short Term'), and occurred with high prevalence, also in agreement with systematic review findings. The fifth trajectory subgroup, 'Intermittent Low', had a less clearly identifiable pattern other than being episodic fit notes, suggesting a subgroup of individuals who were in-and-out of absence during the one-year follow-up.

The final study (Chapter 8) identified a set of common characteristics found to be associated with longer-term absence trajectories for both the MSK and MH condition fit note cohorts; individuals who were: older, living in the North or Midlands or more deprived areas of England, prescribed opioids in the two years preceding their index fit note, and current smokers.

The face validity of all findings from this thesis were affirmed through discussions with a GP, OHID and a PPIE group.

9.1.2 Key Thesis Findings

In this Section, the key thesis findings from the previous Section are reiterated and summarised in Box 6.

Box 6. Summary of Key Thesis Findings

Key Findings from Thesis:

- There have been few studies of trajectories of sickness absence to date (only six were identified in this thesis' systematic review) and none conducted in UK.

- Using trajectory methodology as applied to work absence research is still a relatively new area (the earliest study identified in the systematic review was conducted in 2016), and absence trajectory methods have generally been reported with low quality.

- Incidence of fit note issuance in England due to a MSK condition decreased from 2014 to 2021, whilst incidence due to a MH condition increased.

- Younger females and people living in the Northeast or Northwest of England were identified as having higher incidence of fit notes due to a MH condition, whilst there were no clear sociodemographic differences due to a MSK condition.

- Five different absence trajectories were identified in this thesis, for an English population with baseline absence due to either a MSK or MH condition over 2016-2018, and with a one-year follow-up based on two-monthly recurring intervals.

- Two trajectories consisting of longer-term absence were identified ('Chronic Sustained' and 'Chronic Fast Decreasing' classes), and in line with the systematic review findings, these slower RTW trajectories occurred with lower prevalence.

- In contrast, two trajectories of shorter-term absence were also identified ('Single' fit note issuance, and 'Short Term' absence), which occurred with greater prevalence (also in line with systematic review findings).

- The fifth trajectory represented a subgroup with intermittent spells of absence during follow-up.

- Finally, trajectory-covariate association analysis revealed subgroups identified as higher risk for long-term absence due to either a MSK or MH condition: older people, living in the North or Midlands or more deprived areas of England, people that were prescribed opioids in the two years prior to incident absence, and current smokers.

Abbreviations: MSK = Musculoskeletal; MH = Mental Health.

9.1.3 Setting Thesis Findings In Context

This Section sets the findings from this thesis into a wider context, especially concerning health inequality in England.

Firstly, the extent of people that end up on long-term sickness absence after an initial absence was quantified in Chapter 7. It was shown that whilst slower RTW trajectories were less common than faster RTW trajectories, 6.3% and 11.1% of economically active individuals in England with a first absence due to a MSK or MH condition, respectively, went on to have long-term absences lasting six months or more from baseline absence. Additionally, NHS Digital data shows that around 50% of all fit notes (not just incident fit notes) issued in England due to a MSK or MH condition in England have a duration of more than one month.[38]

This is concerning, as long-term sickness absence has risen to record numbers and is the main reason for economic inactivity in the UK (currently accounting for 30.2% of the economically inactive population).[1,215] In particular, since the start of the COVID-19 pandemic up to October-December 2023, economic inactivity due to long-term sickness absence has risen by over 500,000 people.[215] This puts the UK economy at a disadvantage compared to that of other Western countries whose economies have since shown recovery towards pre-pandemic levels.[216]

Additionally, being in work is generally good for an individual, whilst being off work can negatively impact their mental and physical wellbeing and cause financial pressures.[3,216] The issue is that the longer an individual is off work, the harder it becomes to RTW.[3]

Research that aims to identify subgroups at risk of long-term or intermittent periods absence, as in this thesis, allows such subgroups to then be targeted for early intervention (the ideal type of intervention required is discussed throughout this Section and the next), which could help promote a quicker and sustained RTW, to help reduce national economic inactivity levels.

In Chapter 8 of this thesis, subgroups of individuals more likely to follow trajectories of longer-term absence were identified, and through this process presence of health inequality was found.

Namely, compared to individuals from the least deprived quintile (20 percent) of neighbourhoods of England, those from the most deprived 20 percent, were more likely to follow one of the two longer-term absence trajectories identified in our study, as opposed to a short-term absence trajectory of a single index fit note. This finding was observed for cohorts with an incident fit note due to either a MSK or MH condition, and for the MH condition cohort there was also presence of a social gradient (where there was increasing association of long-term absence with increasingly deprived quintiles of neighbourhoods).

Additionally, regional inequality was also observed in this thesis, with people living in the North or the Midlands more likely than those living in the South to follow longer-term absence trajectories.

Furthermore, a social gradient of health inequality was observed by age subgroups, with older people more likely to end up on longer-term absence (there was a stronger association for higher age groups: 46-55 or 56-66 years, compared to 16-25 year olds).

Health inequalities in these subgroups of people have been observed for other health outcomes too, including healthy life expectancy, HWLE, working life expectancy and more broadly, life expectancy. For example, Marmot et al showed in the SRHIE (2010) report[18] that those in the most deprived neighbourhoods of England, compared to the least deprived, lived shorter lives and with more disability. Moreover, Parker et al (2020)[21] showed that these people from worse off areas of England also had a lower HWLE than those from better off areas, with implications concerning the upcoming increase in SPA, as this subgroup may not have the financial means to take an early retirement and hence might be more disadvantaged by such policy changes.

The factors that influence long-term sickness absence inequalities in neighbourhood deprivation, geographical region and age, may be the same factors influencing inequalities in these subgroups across other health outcomes. Indeed, Marmot et al (2010, 2020)[18,22] posit that the social factors affecting all health inequalities are interconnected, hence to reduce health inequalities completely is a difficult and complex problem that requires policy intervention against all of these social determinants. These factors are wide in scope, and include differences in: housing and neighbourhood conditions, early child development and education, type of employment and working conditions.

Addressing the social determinants of health inequalities remains a pressing issue today. The latest report by the UCL Institute of Health Equity: Health Inequalities, Lives Cut Short (2024),[217] builds upon the SRHIE (2010)[18] and 'The Marmot Review 10 Years On' (2020),[22] and shows that health inequalities in England have continued to widen due to the cumulative impact of budget cuts during the periods of austerity in the UK from 2010 that followed the Global Financial Crisis, the COVID-19 pandemic, and the

cost-of-living crisis. For example, from 2011 to the start of the COVID-19 pandemic over a million people died prematurely, compared to the life expectancy of people living in the least deprived decile of English neighbourhoods.[217]

Key to addressing health inequality is a joined-up approach involving action from all stakeholders concerned, including central and local Government authorities and the third and private sectors.[18,22] This is important, as action from only the NHS and Department of Health is not enough. Indeed, evidence shows that access to health care only accounts for 20% of population health; the majority of our health is influenced by social factors.[217] In a sickness absence context, evidence has shown that treatment for health conditions alone, whilst important, only plays a small role by itself in helping people RTW.[34] Rather, work-focused healthcare, administered in combination with an accommodating employer (a key stakeholder in the RTW process) and through addressing wider social barriers preventing a RTW has been shown to be effective in achieving RTW.[34]

In other words, taking into account the wide range of social determinants of health inequalities mentioned earlier, a holistic and multistakeholder approach is required in addressing long-term sickness absence (this is discussed further in the next Section). A personalised intervention is advocated, that considers the individual as a whole, rather than the emphasis only being on treating the health condition that causes their absence.

Nonetheless, in Chapter 8 of this thesis, it was shown that severity of health conditions did have some influence on the likelihood of following a longer-term sickness absence trajectory. For those who were off work for the first time due to a MSK condition, they were more likely to end up on longer-term absence if they had more (≥2) MSK consultations in the two years prior to their index fit note, and if they were prescribed

any of the following drugs in this time period: opioids, gabapentinoids or antidepressants. Number of prior MSK consultations and drug issuance was used as a marker for severity of MSK conditions in this thesis, as direct measures of severity were not available.

For a cohort of people off work for the first time due to a MH condition, the risk of ending up on a long-term absence trajectory was increased if the baseline MH condition was anxiety, depression, or anxiety and depression, compared to stress. Being prescribed opioids was also associated with longer-term absence, which may be a marker for more severe mental ill-health.

Another key finding from this thesis for people with a first fit note due to a MH condition was presence of sex inequality, with males having an increased risk of long-term absence compared to females. In contrast to this finding, in Chapter 6, females (aged 16-25 years old) were identified as having a higher risk of issuance of a first ever fit note for a MH condition. Yet, when examining absence behaviour over follow-up, (younger) females were generally more likely to follow a trajectory characterised by issuance of only a single index fit note during a one-year follow-up, rather than longer-term absence (Chapter 8). This difference in findings might be explained by females seeking the help that they need to manage their MH conditions earlier in their working lives, which may then reduce the impact of these conditions later on and thereby reduce the risk of this subgroup falling into a spiral of repeated sickness absence episodes.

Finally, further findings from this thesis relate to lifestyle factors. Smokers, compared to people that had never smoked, were more likely to experience long-term sickness absence trajectories for people who had a first fit note due to either a MSK or MH condition. Additionally, in the cohort of people with a first fit note due to a MH

condition, people who were obese, compared to people with an underweight or normal BMI, were more likely to experience longer-term sickness absence. These findings further add to the need for a wide range of interventions when tackling long-term absence, to address smoking and obesity. For example, interventions such as promotion of healthier eating habits and lifestyles, including opportunity for exercise, and smoking cessation programs could be beneficial. Indeed, research has shown that interventions to improve nutrition and fitness in the workplace can have a positive impact on work outcomes such as absenteeism.[218]

To conclude, in this Section the key findings of this thesis were presented in relation to the quantification of long-term sickness absence, as well as the wider societal impact of sickness absence health inequalities for subgroups of people living in more deprived neighbourhoods or the North of England or Midlands, and in older people. The need for a holistic and multistakeholder approach to intervention for reducing long-term sickness absence was discussed, that includes not only the treatment of a health condition, but also the many social factors affecting sickness absence inequalities. In the next Section, an important upcoming pilot scheme is discussed that aims to provide such a holistic and multistakeholder approach to addressing Britain's missing worker problem: WorkWell.[216]

9.1.4 Policies to Reduce Long-Term Sickness Absence

In this Section, options to reduce long-term sickness absence in England are discussed.

Firstly, the main focus of this Section is on WorkWell, a novel pilot scheme from the DWP and DHSC as part of the Government's response to tackling the rise of economic inactivity due to long-term sickness absence in England.[216] WorkWell is a low-intensity,

holistic early intervention of work and health support designed to help people RTW, stay in work (for those who are in work but at risk of falling out), and to help those who are unemployed to find work.

WorkWell services will be available for up to 59,000 disabled people or people with health conditions (with a focus on MSK and MH conditions), with around £57 million of funds available (through successful grant application) for up to 15 different areas across England. Integrated Care Systems (ICSs) can apply for WorkWell funding – ICSs are partnerships that were formed on 1st July 2022 and cover all areas of England, they aim to reduce health inequalities and improve health outcomes through a joined-up approach between local councils, the NHS, and the voluntary sector, amongst other organisations.[219]

Application of WorkWell to a local area is permitted to be flexible (i.e., WorkWell is not a fixed one-size-fits-all approach to reducing long-term sickness absence). ICSs can choose to implement the scheme according to the needs of the people in their locality as well as the local services available, as long the application of WorkWell adheres to the general guidelines set. Delivery of WorkWell is expected to begin in Autumn 2024 and last for two years, with the possibility of rolling this service out nationally if the pilot schemes are deemed to be successful (an external evaluation of the programme is planned).

Once a person is enrolled onto WorkWell, they will be assigned a work and health coach who will be their first point of contact and will perform a holistic assessment of their barriers to RTW. This work and health coach will work with the individual to set up a clear RTW plan that addresses their personal biopsychosocial needs. This plan can involve employer liaison (if the individual consents) to incorporate employer feedback,

and so that the employer can, for example, assist with implementing workplace adjustments etc. Progress towards a RTW will be monitored through ongoing support sessions as required.

Furthermore, WorkWell will function as a signposting service for individuals. The work and health coach will be trained to refer the individual onwards for not only clinical support to a HCP for further treatment, but also for more holistic support which can be as wide ranging as: debt advice, talking therapy, or engagement with any local community or council services that may help the individual, such as a running, cooking, gardening, or singing club/activity. The fundamental aim is to use the local knowledge of the services available in the area to the maximum benefit for the absence recovery of an individual. Knowledge sharing of what works between different WorkWell sites is also encouraged.

Indeed, the core of WorkWell is a joined-up service between multiple stakeholders within ICSs, including but not limited to: employers, HCPs, social workers, charities, voluntary and community services, health promotion programs, and financial support services.

All of the above elements of delivery of WorkWell mentioned in this Section thus far will vary according to each individual ICS.

Additionally, it is planned for some of the WorkWell participating ICSs to help run small-scale pilots of a different Governmental policy that is currently being discussed in an effort to help reduce long-term sickness absence: a reform of the fit note.[216,220] This potential reform would enable people who are issued long-term fit notes to receive

access to specialised health and work support quicker, through better signposting and triaging; further details have yet to be announced.[220]

The relevance of WorkWell to this thesis is clear, WorkWell aligns well with this thesis' aims (Chapter 2) as part of WorkWell's focus is on providing an early intervention for individuals at risk of a long-term sickness absence due to a MSK or MH condition in England. Furthermore, WorkWell provides the type of holistic and multistakeholder approach that was advocated in the previous Section. The profiles of individuals uncovered in Chapter 8 of this thesis, of people at risk of an intermittent or long-term sickness absence following a first ever fit note due to a MSK or MH condition could be used to help identify individuals for early referral to WorkWell, at the point that they receive that first fit note from their HCP. It is hoped that the WorkWell service may then help these individuals achieve a sustained RTW quicker than they would have without this support.

In practice, more may be needed to tackle rising long-term sickness absence in England than the deployment of the WorkWell service alone. As mentioned in the previous Section, eliminating health inequality entirely is an ambitious and complex task, requiring action against all social determinants of health inequality. For example, improving opportunities in the jobs market, especially in more deprived neighbourhoods of England may be an important action in reducing health inequalities based on neighbourhood deprivation.[18] Additionally, funding for schemes to promote a healthier nation may be needed, especially schemes that provide equal opportunity to all, such as: investment for more green spaces in cities, more cycling lanes, and cheaper/subsidised public transport for those who cannot afford it.[22]

Furthermore, policy to incentivise employers for engaging in training concerning RTW management of their employees may also be beneficial, to help employers become more accommodating and supportive of their employees' needs. Employer support arose as a key theme that a PPIE group identified as being important to help them RTW quicker – this is expanded upon further in Section 9.2.1.

In conclusion, a holistic and multistakeholder health and work support intervention for those identified as at risk of an intermittent or long-term sickness absence, alongside broader national policies to improve the health of the nation and reduce health inequality, may help to reduce the number of people on long-term sickness absence.

9.2 Thesis Strengths and Limitations

In this Section, the main strengths and limitations across the studies conducted in this thesis are considered as a whole and discussed.

9.2.1 Generalisability and Validity of Findings

Firstly, by using EHR data of fit note issuance from CPRD Aurum throughout this thesis, one of the main strengths of all studies conducted was that the findings are expected to be generalisable to the wider English population (Wolf et al[128] have shown that CPRD Aurum is representative of the general English population with respect to age, gender, deprivation and geographical region).

Furthermore, another principal strength of this thesis concerned the large sample size in all studies, obtained through using CPRD Aurum. When analysing yearly patterns of first ever fit note issuance in Chapter 6, n=375,113 incident fit notes due to a MSK or MH condition were analysed in total. This then allowed a large sample to be used in Chapter 7, when trajectories of absence were derived using a subset of the study

population from Chapter 6. The final trajectory models that were chosen contained n=43,130 and n=62,355 individuals, for cohorts with a first fit note due to a MSK or MH condition, respectively.

As demonstrated in Section 7.2.2, these large sample sizes from CPRD Aurum allowed the trajectory derivation study to be highly powered for detecting trajectory prevalences. This high statistical power is especially important for trajectory classes that occur less often, as it reduces the risk of spurious findings that might be specific to the dataset being analysed, and not necessarily replicable in other datasets. During the model building strategy, it was initially planned to use suggested guidelines for well performing models and to require a minimum trajectory prevalence of ≥ 5% in all classes,[170,171] however, the high statistical power available through CPRD Aurum allowed this requirement to be relaxed to ≥ 1% (see Section 5.2.6 for more detailed explanation). In the cohort of people with a first fit note due to a MSK condition, the chosen optimal trajectory model did indeed contain two trajectory classes with prevalence < 5% ('Chronic Sustained' and 'Chronic Fast Decreasing', with prevalences of 3.7% and 2.6%, respectively); these classes might not have been identified if a smaller dataset had been used.

Additionally, a key strength of this thesis was that it built upon the scarce evidence base of absence trajectory research identified from the systematic review performed in Chapter 3, and was the first study to examine trajectories of absence for an English population. As explained in Sections 9.1.3 and 9.1.4, this is important, as economic inactivity due to long-term sickness absence is rising in England, and by identifying trajectories of absence, subgroups with higher risk of intermittent or longer-term

absence can then be uncovered and could subsequently be prioritised for a health and work intervention.

Novel statistical methodology taking an individual-centred approach to longitudinal analysis of fit note issuance patterns, using types of LCA was used in this thesis. Furthermore, building on from the critical appraisal of trajectory derivation methods of included studies in the systematic review (Section 5.4), which showed that absence trajectory methodology was generally poorly reported, another strength of this thesis was the high quality trajectory reporting throughout, performed in accordance with the GRoLTS checklist[173] (Appendix G). This is important, as high-quality reporting facilitates critical appraisal of the analysis used (which was challenging to do for the studies in the systematic review), as well as transparency and reproducibility of the analyses conducted.[62,173]

Finally, another key strength of this thesis, was that face validity of the findings from all three studies were affirmed through a diverse group of stakeholders: a GP, OHID, and a PPIE group.

In particular, in the PPIE group session that AL led on 1st February 2024, members of the public with lived experience of sickness absence due to a MSK or MH condition added valuable insight to this thesis' findings. A key theme that emerged was the importance of the role that employers play in encouraging a RTW. Members shared personal experiences about the challenges in trying to RTW after a lengthy absence, and how they can feel "useless" upon returning and can wonder "why have I come back?". One member said, "if you have an understanding employer, it makes such a difference." The impact of small things that the employer can do was discussed too, such as meeting a returning employee in the car park on their first day back and walking them into the

office. Members were in agreement that a policy to provide incentives to employers who provide RTW support (or a penalty for employers that do not) could be beneficial to quicken the recovery process during absence. The general consensus was that more awareness for employers was needed in how they can facilitate a RTW in their employees.

In the PPIE session, the value of a more holistic approach, rather than relying on pharmaceutical treatments, was discussed as being useful too, especially for those following longer-term absence trajectories. Better signposting and referrals from the HCP during a consultation for a fit note was advocated, with one member saying, "there's never that synergy of treating the whole person, rather than treating the symptoms." This ties in with the type of holistic service that WorkWell plans to offer, as discussed in Section 9.1.4.

To conclude, in this Section the general strengths of this thesis were discussed, especially in relation to using CPRD Aurum for analyses. However, using CPRD Aurum also resulted in some study limitations, which are discussed in Sections 9.2.2 and 9.2.3.

9.2.2 Continuous Fit Note Duration Data

One of the key limitations of using CPRD Aurum in this thesis was that fit note duration data was not available for use in analysis. Less than 0.5% of individuals with a first ever fit note due to either a MSK or MH condition during 2010-2021 from study 1 had fit note duration data present.

It transpired that fit note data was largely recorded as missing in the CPRD Aurum database. AL investigated the cause of this issue with a GP and a Health Informatics

Specialist from Keele University, and found that GPs do indeed record fit note duration after issuing a fit note by entering a numerical duration value (in days or weeks) or by specifying a fit note end date (entering a value in either the duration or fit note end date fields is a mandatory part of the data-entry process). However, this fit note duration/end date data is automatically converted into free-text format after the GP has completed recording of the issued fit note. The problem is then that CPRD do not accept any free-text data, for reasons of risk of patient identification.[221] The Data Manager of Medical Records at Keele University, JB, contacted CPRD who were not aware of any means to get access to this duration data through their databases.

Therefore, an alternative potential source of fit note duration data was considered, through NHS Digital, who have rich data regarding fit notes issued in England. NHS Digital produce quarterly reports of fit note issuance which includes summaries of fit note duration by issuance month and reason for fit note. AL contacted NHS Digital on 28th October 2022 and enquired whether it would be possible to request individual-level data concerning fit note duration through their Data Access Request Service, but was informed that it was not possible to apply for access to any of their fit note data, and that there was no timescale on when this might become possible.

Best efforts were then made by AL to investigate the possibility of inferring fit note duration data from the data that was available in CPRD Aurum (through using start dates of fit notes for a given individual, and making an assumption that the start date of a second fit note for a given individual was the end date of the first fit note etc, as long as the gap between two successive fit notes was less than six months). However, as explained in Section 7.2.3, this approach was ultimately not feasible.

As a result, trajectory derivation analysis in this thesis was conducted using a binary fit note definition, measured as 1/0 for fit note issuance in a given interval. Whilst the face validity of the five optimal absence trajectories uncovered in this thesis was confirmed through discussion with stakeholders (discussed in previous Section), and similarities were seen between the fastest and slowest RTW trajectories in the absence trajectory studies included in this thesis' systematic review, it is possible that even more nuanced patterns of absence over time might have been uncovered had a continuous fit note definition been used.

The potential advantages of a continuous fit note definition are explained throughout the rest of this Section.

Firstly, by using a fit note definition based on whether or not a fit note was issued in a given interval, the minimum length of time intervals explored for operational performance in the trajectory derivation was limited to two months in this thesis. The absence trajectory studies in this thesis' systematic review all used either weekly or monthly time intervals for trajectory derivation, but it was not possible to use these intervals in this thesis as around 50% of fit notes issued due to a MSK or MH condition in England have been shown to have duration > one month,[38] hence sickness absence could have been underestimated if smaller time intervals had been used. It was shown in Chapter 7 that by reducing time intervals in trajectory derivation models from six to three to two months, the performance of the models improved progressively. Therefore, it is plausible that using monthly time intervals may have further improved model performance.

Using a continuous definition of fit note issuance with monthly intervals, may have also led to more complex trajectory functional forms being uncovered in the optimal models

(rather than linear, as in this study). This is because, firstly, by using intervals of shorter time periods, the number of repeated measurements being analysed is increased. Secondly, use of a continuous repeated measure definition as opposed to a binary definition allows for more possible options in a given time interval. For example, a binary repeated measure only provides two options for an individual in a given time interval (such as 1/0 for fit note issuance), whereas if a continuous measure is used based on average days of absence in a 30-day period, each individual has up to 30 possible options (assuming absence days are recorded as whole days and not partial absence days). Thus, through analysing an increased number of repeated measurements and with there being an increased number of possible options at each measurement, the complexity of possible patterns that can be modelled with a continuous fit note definition grows exponentially.

More complex optimal trajectory shapes could have led to better fitting models and perhaps uplifted the entropy values of this thesis' optimal models from 0.72 and 0.73 (for the index MSK and MH condition index fit note cohorts, respectively), to values more in line with Rysstad et al's[101] optimal model that had a higher entropy of 0.95 (this study was based on a continuous definition of absence with monthly intervals, and featured quadratic and cubic optimal trajectories).

However, a limitation of using a continuous fit note definition is that models with an increased complexity may be more difficult to interpret, and therefore such models may be less practical to use. For example, engaging with stakeholders to explain the findings and seek their feedback might be more challenging if stakeholders do not have a technical statistical background. Furthermore, more complex models may have more convergence issues, or may not run at all. In this instance, a binary or categorical fit

note definition may need to be used, such as through using categories of fit note duration (e.g., 0 days, 1-7 days, 8-14 days, etc) rather than the actual values.

In summary, it is possible that using a continuous fit note definition for trajectory derivation may lead to trajectory models with improved statistical performance to those of this thesis, and possibly uncover more nuanced patterns of absence. However, as the optimal trajectories identified in this thesis based on a binary fit note definition had face validity, were considered plausible, and were generally similar to those from the studies in this thesis' systematic review, they are considered to reflect the aims of this thesis well and replication of analysis based on a continuous fit note definition is not deemed a priority for future work.

9.2.3 Further Covariates for Trajectory-Covariate Association Analysis

To conclude this Section, another key limitation with using CPRD Aurum relates to the scope of characteristics that were available for exploration in trajectory-covariate association analysis (Chapter 8).

Though a broad range of characteristics were tested for association with optimal sickness absence trajectories in Chapter 8 - sixteen characteristics in total, encompassing different sociodemographic and health characteristics, types of treatment received, and comorbidity, some potentially important covariates were not tested as they are not available from primary care records.

For example, a 2016 synthesis of systematic reviews of prognostic factors for RTW for absence due to a range of health conditions (including MSK and MH conditions), found that important common factors related to faster RTW included: higher education and socioeconomic status, higher self-efficacy to RTW, lower severity of health condition,

and workplace factors such as RTW coordination, whereas examples of factors associated with slower RTW included: prior sick leave and unemployment, and more physically demanding jobs.[222]

In this thesis, deprivation was used as a proxy for socioeconomic status/educational level, and prior MSK/MH condition consultations and drug prescriptions as a marker for severity of health condition. However, the key characteristics that were relevant for the aims of this thesis but not available through CPRD Aurum data linkages, were self-efficacy to RTW and work-related characteristics.

In particular, self-efficacy to RTW is important as individuals' recovery expectations have been identified in other studies as a strong predictor of RTW.[211–214] Negative RTW expectancy was one of the most important predictors discovered in the study by Rysstad et al,[101] that was associated with all three of their trajectories that involved the most sickness absence days.

Furthermore, it would be useful to assess the impact of the following work-related characteristics on association with longer-term absence trajectories: full time compared to part-time work, type of work (unskilled, manual, skilled etc), level of job control, emotional demands, interpersonal relations, amongst others.

Having access to such data might further identify specific subgroups that are at greater risk of intermittent or long-term absence, and that could benefit from early and more targeted health and work support after onset of an initial sickness absence.

The issue is that self-efficacy to RTW and the types of work data described linked to fit notes are not readily available in UK research databases, thus this remains a study limitation. Access to this data could be achieved through a method such as primary data

collection, through conducting a prospective cohort study where this information is obtained by sending a questionnaire to study participants. However, this would be more costly than using secondary data collection as in this thesis, and for the aims of this thesis, where trajectory derivation is required, a large sample size would be needed to provide sufficient power for such complex longitudinal analyses which would further exacerbate costs.

Other limitations of this thesis were discussed separately in the Discussion subsections of earlier Chapters. For example, limitations concerning the definition of a MSK or MH condition fit note used in this thesis, and the assumptions behind this definition, were discussed in detail in Section 6.4.2.

9.3 Further Research

The final Section of this Chapter contains recommendations for further research.

Firstly, though the optimal trajectory models of this thesis generally performed well within the CPRD Aurum data used, and similarity of trajectories was noted against the studies from this thesis' review of absence trajectories, further external validation could be performed.

For example, if it becomes possible to access fit note data through NHS Digital in the future, this could be a rich data source to use for testing external reliability.

Another possible dataset to consider for external validation, and which would also address another limitation of this thesis – that all analyses were restricted to an English population only, and not the other devolved nations of the UK - is CPRD GOLD. CPRD GOLD contains few currently registered patients from England (1.9% from the December 2023 CPRD GOLD monthly release (https://doi.org/10.48329/30pm-

xq61)),[132] and a large proportion from Scotland (56.8%). However, a limitation of this approach is that linkage data on deprivation would not be available other than for patients in England when using CPRD GOLD (the CPRD deprivation linkage is currently only available for English GP practices), which would restrict replication of trajectory-covariate association analysis.

External validation could also be performed using SAIL Databank,[114] to test whether the absence trajectories identified in this thesis hold in a Welsh population. SAIL Databank[114] has a wide range of data linkages available, including access to the Labour Force Survey which contains some data relating to employment.[223]

Although external validation could also be extended more widely, to datasets in other countries outside of the UK, the difference in absence management systems between the UK and other countries would be a hindrance for such analyses, hence this is not recommended.

Additionally, further research could also assess whether using different trajectory derivation methods (aside from LCGA and GMM), such as sequence analysis, results in similar absence trajectories to those uncovered in this thesis.

Furthermore, though the PPIE group session that was conducted on 1st February 2024 by AL provided some level of insight into what kind of support those at risk of intermittent or long-term sickness absence might benefit most from, it is recommended that this work is followed up using a more extensive approach than a PPIE meeting alone. It would be useful to conduct a qualitative study that specifically follows up the subgroups of individuals identified in this thesis as high risk for intermittent or long-term sickness absence due to a MSK or MH condition, and to interview such people

during follow-up with the aim to understand what barriers these people face to RTW, and what support could help them reach a sustained RTW more quickly. Such qualitative research could be used alongside learnings from the upcoming WorkWell pilot (Section 9.1.4), to help shape development of an appropriate work and health intervention.

9.4 Final Conclusions

This thesis has highlighted different patterns of sickness absence in an English population due to a MSK or MH condition and profiles of individuals associated with intermittent and longer-term absence. Face validity of the findings from this thesis have been confirmed through discussion with a GP, OHID and a PPIE group. Furthermore, the optimal derived trajectories showed similarity with those from absence trajectory studies found in this thesis' systematic review.

An earlier and more targeted multistakeholder health and work intervention, based on a holistic approach towards individuals identified in subgroups at higher risk of an intermittent or long-term sickness absence due to a MSK or MH condition, alongside wider national policies to reduce health inequalities, may help alleviate Britain's current missing worker problem.

References

1. Office for National Statistics (ONS). *Half a Million More People Are out of the Labour Force Because of Long-Term Sickness*.; 2022.
2. House of Lords; Economic Affairs Committee. *Where Have All the Workers Gone?*; 2022.
3. Waddell G, Burton KA. *Is Work Good for Your Health and Well-Being?*; 2006.
4. Scheid TL, Anderson C. Living with chronic mental illness: Understanding the role of work. *Community Ment Health J*. 1995;31(2):163-176.
5. UK Government. Statutory Sick Pay (SSP). Accessed October 20, 2023. https://www.gov.uk/statutory-sick-pay
6. Office for National Statistics (ONS). *Employee Earnings in the UK: 2023*.; 2023.
7. Virtanen P, Liukkonen V, Vahtera J, Kivimäki M, Koskenvuo M. Health inequalities in the workforce: the labour market core–periphery structure. *Int J Epidemiol*. 2003;32(6):1015-1021.
8. Wilkinson RD, Pickett K. *The Spirit Level: Why More Equal Societies Almost Always Do Better*. London: Allen Lane; 2009.
9. Wynne-Jones G, Cowen J, Jordan JL, Uthman O, Main CJ, Glozier N. Absence from work and return to work in people with back pain: A systematic review and meta-analysis. *Occup Environ Med*. 2014;71(6):448-458.
10. NICE. NICE and health inequalities. Accessed January 7, 2024. https://www.nice.org.uk/about/what-we-do/nice-and-health-inequalities
11. Public Health England. *Guidance: Place-Based Approaches for Reducing Health Inequalities: Main Report*.; 2021.
12. WHO CSDH. *Closing the Gap in a Generation: Health Equity through Action on the Social Determinants of Health. Final Report of the Commission on Social Determinants of Health*.; 2008.
13. Department of Health and Social Security. *Inequalities in Health: Report of a Research Working Group*. HM Stationery Office; 1980.
14. Gray AM. Inequalities in Health. The Black Report: A Summary and Comment. *Int J Heal Serv*. 1982;12(3):349-380.
15. Marmot MG, Stansfeld S, Patel C, et al. Health inequalities among British civil servants: the Whitehall II study. *Lancet*. 1991;337(8754):1387-1393.
16. Marmot MG, Rose G, Shipley M, Hamilton PJ. Employment grade and coronary heart disease in British civil servants. *J Epidemiol Community Health*. 1978;32(4):244-249.
17. North F, Syme SL, Feeney A, Head J, Shipley MJ, Marmot MG. Explaining socioeconomic differences in sickness absence: the Whitehall II Study. *BMJ*.

1993;306(6874):361-366.

18. Marmot M, Allen J, Goldblatt P, et al. *Fair Society, Healthy Lives - The Marmot Review: Strategic Review of Health Inequalities in England Post-2010*.; 2010.

19. Frontier Economics Ltd. Estimating the costs of health inequalities. A report prepared for the Marmot Review. Published 2010. Accessed July 25, 2021. https://www.instituteofhealthequity.org/file-manager/FSHLrelateddocs/overall-costs-fshl.pdf

20. Department for Work and Pensions. Proposed new timetable for State Pension age increases. Published 2017. Accessed August 8, 2021. https://www.gov.uk/government/news/proposed-new-timetable-for-state-pension-age-increases

21. Parker M, Bucknall M, Jagger C, Wilkie R. Population-based estimates of healthy working life expectancy in England at age 50 years: analysis of data from the English Longitudinal Study of Ageing. *Lancet Public Heal*. 2020;5(7):e395-e403.

22. Marmot M, Allen J, Boyce T, Goldblatt P, Morrison J. *Health Equity in England: The Marmot Review 10 Years On*.; 2020.

23. Marmot M. Health equity in England: the Marmot review 10 years on. *BMJ*. 2020;368:m693.

24. PHE. Public Health Profiles. Accessed July 26, 2021. https://fingertips.phe.org.uk

25. Hiam L, Dorling D, McKee M. Things Fall Apart: the British Health Crisis 2010–2020. *Br Med Bull*. 2020;133(1):4-15.

26. Marmot M, Allen J, Goldblatt P, Herd E, Morrison J. *Build Back Fairer: The COVID-19 Marmot Review. The Pandemic, Socioeconomic and Health Inequalities in England.*; 2020.

27. Roy MJ, McHugh N, Hill O'Connor C. Social Innovation: Worklessness, Welfare and Well-being. *Soc Policy Soc*. 2014;13(3):457-467.

28. Bambra C. *Work, Worklessness, and the Political Economy of Health*. OUP Oxford; 2011.

29. Kasl S V, Jones BA. The impact of job loss and retirement on health. In: *Social Epidemiology*. Oxford University Press New York; 2000:118-136.

30. Bartley M. *Health Inequality: An Introduction to Concepts, Theories and Methods*. 2nd ed. John Wiley & Sons; 2016.

31. Thomas C, Benzeval M, Stansfeld SA. Employment transitions and mental health: An analysis from the British household panel survey. *J Epidemiol Community Health*. 2005;59(3):243-249.

32. Jusot F, Khlat M, Rochereau T, Serme C. Job loss from poor health, smoking and obesity: a national prospective survey in France. *J Epidemiol Community Health*. 2008;62(4):332-337.

33. Department for Work and Pensions. Guidance for GPs, employers, hospital doctors and occupational health practitioners on using fit notes (statements of fitness for work). Published 2018. Accessed August 5, 2021. https://www.gov.uk/government/collections/fit-note

34. Waddell G, Burton KA, Kendall NA. *Vocational Rehabilitation, What Works, for Whom and When?*; 2008.

35. Department for Work and Pensions. *Getting the Most out of the Fit Note: GP Guidance*.; 2015.

36. Welsh VK, Mallen CD, Wynne-Jones G, Jinks C. Exploration of GPs' views and use of the fit note: a qualitative study in primary care. *Br J Gen Pract*. 2012;62(598):e363-e370.

37. Department for Work and Pensions. *The Fit Note: A Guide for Patients and Employees*.; 2015.

38. NHS Digital. *Fit Notes Issued by GP Practices, England, September 2023*.; 2024.

39. Black C. *Working for a Healthier Tomorrow*.; 2008.

40. Institution of Occupational Safety and Health. *Reforming the Medical Statement - Consultation on Draft Regulations, May 2009*.; 2009.

41. Black C, Frost D. *Health at Work – an Independent Review of Sickness Absence*.; 2011.

42. Yarker J, Munir F, Donaldson-Feilder E, Hicks B. *Managing Rehabilitation: A Competency Framework for Managers to Support Return to Work*.; 2010.

43. Thomson L, Rick J. An organisational approach to the rehabilitation of employees following stress-related illness. In: Kinder A, Hughes R, Cooper C, eds. *Employee Well-Being Support: A Workplace Resource*. Wiley & Sons; 2008.

44. Franche R-L, Cullen K, Clarke J, Irvin E, Sinclair S, Frank J. Workplace-based return-to-work interventions: a systematic review of the quantitative literature. *J Occup Rehabil*. 2005;15(4):607-631.

45. Hensing G, Alexanderson K, Allebeck P, Bjurulf P. How to measure sickness absence? Literature review and suggestion of five basic measures. *Scand J Soc Med*. 1998;26(2):133-144.

46. Hensing G. Chapter 4. Methodological aspects in sickness-absence research. *Scand J Public Health*. 2004;32:44-48.

47. NICE. Workplace health: long-term sickness absence and capability to work. NICE Public Health Guideline NG146. Published 2019. Accessed July 25, 2021. https://www.nice.org.uk/guidance/ng146

48. Thomson L, Hampton R. Fit for work? Changing fit note practice among GPs. *Br J Gen Pract*. 2012;62(595):e147-e150.

49. Hann M, Sibbald B. *General Practitioners' Attitudes towards Patients' Health and Work, 2010–12*.; 2013.

50. Dorrington S, Roberts E, Mykletun A, Hatch S, Madan I, Hotopf M. Systematic review of fit note use for workers in the UK. *Occup Environ Med*. 2018;75(7):530-539.

51. DWP. *Improving Lives: The Future of Work, Health and Disability*.; 2017.

52. Kausto J, Solovieva S, Virta LJ, Viikari-Juntura E. Partial sick leave associated with disability pension: propensity score approach in a register-based cohort study. *BMJ Open*. 2012;2(6).

53. Parsonage M, Saini G. *Mental Health at Work: The Business Costs Ten Years On*.; 2017.

54. HSE. *Work Related Musculoskeletal Disorder Statistics (WRMSDs) in Great Britain, 2020*.; 2020.

55. Wynne-Jones G, Dunn KM. Has there been a change in the rates of UK sickness certification for back pain over time? An examination of historical data from 2000 to 2010. *BMJ Open*. 2016;6(4):e009634.

56. Henderson M, Harvey SB, Øverland S, Mykletun A, Hotopf M. Work and common psychiatric disorders. *J R Soc Med*. 2011;104(5):198-207.

57. Wynne-Jones G, Mallen CD, Mottram S, Main CJ, Dunn KM. Identification of UK sickness certification rates, standardised for age and sex. *Br J Gen Pract*. 2009;59(564):510-516.

58. Gabbay M, Shiels C, Hillage J. Sickness certification for common mental disorders and GP return-to-work advice. *Prim Health Care Res Dev*. 2016;17(5):437-447.

59. Gjesdal S, Holmaas TH, Monstad K, Hetlevik Ø. GP consultations for common mental disorders and subsequent sickness certification: register-based study of the employed population in Norway. *Fam Pract*. 2016;33(6):656-662.

60. CPRD. Clinical Practice Research Datalink (CPRD). Published 2024. Accessed January 4, 2024. https://cprd.com/

61. Jung T, Wickrama KAS. An Introduction to Latent Class Growth Analysis and Growth Mixture Modeling. *Soc Personal Psychol Compass*. 2008;2(1):302-317.

62. Nguena Nguefack HL, Pagé MG, Katz J, et al. Trajectory Modelling Techniques Useful to Epidemiological Research: A Comparative Narrative Review of Approaches. *Clin Epidemiol*. 2020;12:1205-1222.

63. Dunn KM, Jordan K, Croft PR. Characterizing the course of low back pain: a latent class analysis. *Am J Epidemiol*. 2006;163(8):754-761.

64. Kongsted A, Kent P, Axen I, Downie AS, Dunn KM. What have we learned from ten years of trajectory research in low back pain? *BMC Musculoskelet Disord*. 2016;17:220.

65. Ogollah RO, Konstantinou K, Stynes S, Dunn KM. Determining One-Year Trajectories of Low-Back-Related Leg Pain in Primary Care Patients: Growth Mixture Modeling of a Prospective Cohort Study. *Arthritis Care Res (Hoboken)*.

2018;70(12):1840-1848.

66. Nicholls E, Thomas E, van der Windt DA, Croft PR, Peat G. Pain trajectory groups in persons with, or at high risk of, knee osteoarthritis: findings from the Knee Clinical Assessment Study and the Osteoarthritis Initiative. *Osteoarthr Cartil*. 2014;22(12):2041-2050.

67. Tran T, Hammarberg K, Ryan J, et al. Mental health trajectories among women in Australia as they age. *Aging Ment Health*. 2019;23(7):887-896.

68. Saunders R, Buckman JEJ, Fonagy P, Fancourt D. Understanding different trajectories of mental health across the general population during the COVID-19 pandemic. *Psychol Med*. Published online 2021:1-9.

69. Farrants K, Friberg E, Sjolund S, Alexanderson K. Trajectories of future sickness absence and disability pension days among individuals with a new sickness absence spell due to osteoarthritis diagnosis >=21 days: a prospective cohort study with 13-month follow-up. *BMJ Open*. 2019;9(8):e030054.

70. McLeod CB, Reiff E, Maas E, Bultmann U. Identifying return-to-work trajectories using sequence analysis in a cohort of workers with work-related musculoskeletal disorders. *Scand J Work Environ Health*. 2018;44(2):147-155.

71. Feleus A, Miedema HS, Bierma-Zeinstra SMA, Hoekstra T, Koes BW, Burdorf A. Sick leave in workers with arm, neck and/or shoulder complaints; defining occurrence and discriminative trajectories over a 2-year time period. *Occup Environ Med*. 2017;74(2):114-122.

72. Farrants K, Friberg E, Sjölund S, Alexanderson K. Work disability trajectories among individuals with a sick-leave spell due to depressive episode ≥ 21 days: A prospective cohort study with 13-month follow up. *J Occup Rehabil*. 2018;28(4):678-690.

73. Spronken M, Brouwers EPM, Vermunt JK, et al. Identifying return to work trajectories among employees on sick leave due to mental health problems using latent class transition analysis. *BMJ Open*. 2020;10(2):e032016.

74. Gopalakrishnan S, Ganeshkumar P. Systematic Reviews and Meta-analysis: Understanding the Best Evidence in Primary Healthcare. *J Fam Med Prim care*. 2013;2(1):9-14.

75. Moher D, Shamseer L, Clarke M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev*. 2015;4(1):1.

76. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan—a web and mobile app for systematic reviews. *Syst Rev*. 2016;5(1):210.

77. Haukka E, Kaila-Kangas L, Ojajärvi A, et al. Pain in multiple sites and sickness absence trajectories: a prospective study among Finns. *Pain*. 2013;154(2):306-312.

78. Dorrington S, Carr E, Polling C, et al. Health condition at first fit note and number of fit notes: a longitudinal study of primary care records in south

London. *BMJ Open*. 2021;11(3):e043889.

79. McKenzie J, Brennan S, Ryan R, Thomson H, Johnston R, Thomas J. Defining the criteria for including studies and how they will be grouped for the synthesis. In: Higgins J, Thomas J, Chandler J, et al., eds. *Cochrane Handbook for Systematic Reviews of Interventions Version 6.2 (Updated February 2021)*. Cochrane; 2021.

80. Wells G, Shea B, O'Connell D, et al. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. Published 2013. http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp

81. CASP. CASP Checklists. https://casp-uk.net/casp-tools-checklists/

82. Hayden JA, van der Windt DA, Cartwright JL, Côté P, Bombardier C. Assessing bias in studies of prognostic factors. *Ann Intern Med*. 2013;158(4):280-286.

83. Ma L-L, Wang Y-Y, Yang Z-H, Huang D, Weng H, Zeng X-T. Methodological quality (risk of bias) assessment tools for primary and secondary medical studies: what are they and which is better? *Mil Med Res*. 2020;7(1):7.

84. Zeng X, Zhang Y, Kwong JSW, et al. The methodological quality assessment tools for preclinical and clinical studies, systematic review and meta-analysis, and clinical practice guideline: a systematic review. *J Evid Based Med*. 2015;8(1):2-10.

85. Nagin DS, Odgers CL. Group-based trajectory modeling in clinical research. *Annu Rev Clin Psychol*. 2010;6:109-138.

86. Barban N, Billari FC. Classifying life course trajectories: a comparison of latent class and sequence analysis. *J R Stat Soc Ser C (Applied Stat*. 2012;61(5):765-784.

87. Cochrane Consumers and Communication Review Group, Ryan R. *Cochrane Consumers and Communication Group: Meta-Analysis*.; 2016.

88. Popay J, Roberts H, Sowden A, et al. *Guidance on the Conduct of Narrative Synthesis in Systematic Reviews: A Product from the ESRC Methods Programme*.; 2006.

89. Bramer W, Bain P. Updating search strategies for systematic reviews using EndNote. *J Med Libr Assoc*. 2017;105(3):285-289.

90. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71.

91. Leao ALM, Barbosa-Branco A, Turchi M, Steenstra I. Recurrence of medically certified sickness absence among newly employed municipal workers. *Occup Environ Med*. 2016;73(Supplement 1):A230.

92. Steenstra I, Hogg-Johnson S, Davilmar A, et al. A prediction rule for duration of disability benefits in workers with nonspecific low back pain. *Occup Environ Med*. 2011;68(SUPPL. 1):A75.

93. Mylle G, Verbeek C, Bulterys S. The impact of the new legislation on return to

work in Belgium. Experiences of a large occupational health service. *Occup Environ Med*. 2018;75(Supplement 2):A540.

94. Hansen SM, Hetland ML, Pedersen J, Ostergaard M, Rubak TS. The prospective risk for long term sickness absence, unemployment, and disability pension, and the probability for return to work in patients with rheumatoid arthritis. *Ann Rheum Dis*. 2016;75(Supplement 2):890.

95. Fishbain DA, Cutler RB, Rosomoff H, et al. "'Movement'" in work status after pain facility treatment. *Spine (Phila Pa 1976)*. 1996;21(22):2662-2669.

96. Booth G, Goldklang G, Roth H. Rehabilitation of indigent arthritis patients in the District of Columbia. *Med Ann Dist Columbia*. 1971;40(5):304-308.

97. Barkley AL, Pixley F, Walker C. Industrial therapy: a study of success rates over an eight-year period. *Rehabil Lit*. 1976;37(5):130-144.

98. Wildenhaus KJ. Feeling the economic effects of back pain. *AHIP Cover*. 2004;45(3):78-84.

99. Segal MS. Back pain: don't overlook evidence. *Minn Med*. 2006;89(3):8-9.

100. Pedersen P, Lund T, Lindholdt L, et al. Labour market trajectories following sickness absence due to self-reported all cause morbidity--a longitudinal study. *BMC Public Health*. 2016;16(1):1-10.

101. Rysstad T, Grotle M, Aasdahl L, Dunn KM, Tveter AT. Identification and Characterisation of Trajectories of Sickness Absence Due to Musculoskeletal Pain: A 1-Year Population-based Study. *J Occup Rehabil*. 2023;33(2):277-287.

102. Borg K, Hensing G, Alexanderson K. Risk factors for disability pension over 11 years in a cohort of young persons initially sick-listed with low back, neck, or shoulder diagnoses. *Scand J Public Health*. 2004;32(4):272-278.

103. Hakulinen C, Bockerman P, Pulkki-Raback L, Virtanen M. Employment and earnings trajectories before and after sickness absence due to major depressive disorder: A nationwide case-control study. *Occup Environ Med*. 2021;78(3):173-178.

104. Ayala-Garcia A, Serra L, Ubalde-Lopez M. Association between early working life patterns, in publicly and privately owned companies, and the course of future sickness absence due to mental disorders: a cohort study in Catalonia (Spain). *BMJ Open*. 2021;11(2):e040480.

105. Hou W, Sheu C, Liang H, Hsieh C, Lee Y, Chuang H. Trajectories and predictors of return to work after traumatic limb injury - a 2-year follow-up study. *Scand J Work Environ Heal*. 2012;38(5):456-466.

106. Bultmann U, Franche R, Hogg-Johnson S, et al. Health status, work limitations, and return-to-work trajectories in injured workers with musculoskeletal disorders. *Qual Life Res*. 2007;16(7):1167-1178.

107. Baldwin ML, Butler RJ. Upper extremity disorders in the workplace: Costs and outcomes beyond the first return to work. *J Occup Rehabil*. 2006;16(3):303-323.

108. Cote P, Baldwin M, Johnson W, Frank J, Butler R. Patterns of sick-leave and health outcomes in injured workers with back pain. *Eur Spine J*. 2008;17(4):484-493.

109. Navarro A, Morina D, Reis R, Nedel FB, Martin M. Hazard functions to describe patterns of new and recurrent sick leave episodes for different diagnoses. *Scand J Work Environ Heal*. 2012;38(5):447-455.

110. Lalic S, Bell JS, Gyllensten H, et al. Trajectories of sickness absence and disability pension before and after opioid initiation for noncancer pain: A 10-year population-based study. *Pain*. 2019;160(5):1224-1233.

111. Wynne-Jones G, Mallen CD, Main CJ, Dunn KM. Sickness certification in general practice: a comparison of electronic records with self-reported work absence. *Prim Health Care Res Dev*. 2008;9(2):113-118.

112. Ghosh RE, Crellin E, Beatty S, Donegan K, Myles P, Williams R. How Clinical Practice Research Datalink data are used to support pharmacovigilance. *Ther Adv drug Saf*. 2019;10:2042098619854010-2042098619854010.

113. Herrett E, Gallagher AM, Bhaskaran K, et al. Data Resource Profile: Clinical Practice Research Datalink (CPRD). *Int J Epidemiol*. 2015;44(3):827-836.

114. SAIL Databank. SAIL DATABANK. Published 2022. Accessed October 26, 2023. https://saildatabank.com/

115. IQVIA. Fact Sheet: UK EMR - IQVIA Medical Research Data. Published 2020. Accessed October 26, 2023. https://www.iqvia.com/library/fact-sheets/uk-emr-iqvia-medical-research-data

116. NHS Health Research Authority. IQVIA Medical Research Data. Published 2018. Accessed October 26, 2023. https://www.hra.nhs.uk/planning-and-improving-research/application-summaries/research-summaries/the-health-improvement-network-thin-database/

117. THIN. The Health Improvement Network. Accessed October 26, 2023. https://www.the-health-improvement-network.com/

118. QResearch. QResearch. Accessed October 26, 2023. https://www.qresearch.org/

119. Mannan F, Chaudhry Z, Gibson-White A, et al. Outputs and growth of primary care databases in the United Kingdom: bibliometric analysis. *BMJ Heal Care Informatics*. 2017;24(3):942.

120. CPRD. Bibliography. Published 2024. Accessed January 4, 2024. https://cprd.com/bibliography

121. Herrett E, Thomas SL, Schoonen WM, Smeeth L, Hall AJ. Validation and validity of diagnoses in the General Practice Research Database: a systematic review. *Br J Clin Pharmacol*. 2010;69(1):4-14.

122. CiPCA Database. Accessed October 26, 2023. https://www.keele.ac.uk/mrr/cipcadatabase/

123. Dorrington S, Carr E, Stevelink SAM, et al. Demographic variation in fit note

receipt and long-term conditions in south London. *Occup Environ Med*. 2020;77(6):418-426.

124. Rathod-Mistry T, Marshall M, Campbell P, et al. Indicators of dementia disease progression in primary care: An electronic health record cohort study. *Eur J Neurol*. 2021;28(5):1499-1510.

125. Mason KJ, Jordan KP, Heron N, et al. Musculoskeletal pain and its impact on prognosis following acute coronary syndrome or stroke: A linked electronic health record cohort study. *Musculoskeletal Care*. 2023;21(3):749-762.

126. NHS Health Research Authority. Clinical Practice Research Datalink (CPRD) Research Database. Published 2022. Accessed January 4, 2024. https://www.hra.nhs.uk/planning-and-improving-research/application-summaries/research-summaries/clinical-practice-research-datalink-cprd-research-database/

127. Kontopantelis E, Stevens RJ, Helms PJ, Edwards D, Doran T, Ashcroft DM. Spatial distribution of clinical computer systems in primary care in England in 2016 and implications for primary care electronic medical record databases: a cross-sectional population study. *BMJ Open*. 2018;8(2):e020738.

128. Wolf A, Dedman D, Campbell J, et al. Data resource profile: Clinical Practice Research Datalink (CPRD) Aurum. *Int J Epidemiol*. 2019;48(6):1740-1740g.

129. CPRD. *CPRD Aurum: Frequently Asked Questions (FAQs)*.; 2021.

130. NHS England. SNOMED CT. https://www.england.nhs.uk/digitaltechnology/digital-primary-care/snomed-ct/

131. NHS Digital. Quality and Outcomes Framework (QOF), enhanced services and core contract extraction specifications (business rules). Published 2022. https://digital.nhs.uk/data-and-information/data-collections-and-data-sets/data-collections/quality-and-outcomes-framework-qof#quality-and-outcome-framework-business-rules

132. CPRD. *CPRD GOLD December 2023 (Version 2023.12.001) [Data Set]. Clinical Practice Research Datalink.*; 2023.

133. CPRD. *CPRD Aurum December 2023 (Version 2023.12.001) [Data Set]. Clinical Practice Research Datalink.*; 2023.

134. CPRD. New Research Data Governance process. Published 2021. https://cprd.com/article/new-research-data-governance-process

135. Kelly MP. *Independent Review of CPRD Internal Data Governance Framework*.; 2019.

136. CPRD. *CPRD Aurum February 2022 (Version 2022.02.001) [Data Set]. Clinical Practice Research Datalink.*; 2022.

137. Karlson B, Jonsson P, Osterberg K. Long-term stability of return to work after a workplace-oriented intervention for patients on sick leave for burnout. *BMC Public Health*. 2014;14:821.

138. Masyn KE, Petras H, Liu W. Growth Curve Models with Categorical Outcomes. In: Bruinsma G, Weisburd D, eds. *Encyclopedia of Criminology and Criminal Justice*. Springer New York; 2013:2013-2025.

139. Stoel RD, van den Wittenboer G, Hox J. Analyzing Longitudinal Data using Multilevel Regression and Latent Growth Curve Analysis. *Metodol las Ciencias del Comport*. 2003;5(1):21-42.

140. Bollen KA. *Structural Equations with Latent Variables.* John Wiley & Sons; 1989.

141. Wickrama K, Lee TK, O'Neal C, Lorenz F. *Higher-Order Growth Curves and Mixture Modeling with Mplus: A Practical Guide*. 2nd ed. Routledge; 2021.

142. McTernan M, Blozis SA. Longitudinal Models for Ordinal Data With Many Zeros and Varying Numbers of Response Categories. *Struct Equ Model A Multidiscip J*. 2015;22(2):216-226.

143. Lee TK, Wickrama K, O'Neal CW. Application of Latent Growth Curve Analysis with Categorical Responses in Social Behavioral Research. *Struct Equ Modeling*. 2018;25(2):294-306.

144. Chen G, Tsurumi H. Probit and Logit Model Selection. *Commun Stat - Theory Methods*. 2010;40(1):159-175.

145. Whittaker TA, Khojasteh J. Detecting Appropriate Trajectories of Growth in Latent Growth Models: The Performance of Information-Based Criteria. *J Exp Educ*. 2017;85(2):215-230.

146. Steiger JH. Structural Model Evaluation and Modification: An Interval Estimation Approach. *Multivariate Behav Res*. 1990;25(2):173-180.

147. Bentler PM. Comparative fit indexes in structural models. *Psychol Bull*. 1990;107(2):238-246.

148. Tucker LR, Lewis C. A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*. 1973;38(1):1-10.

149. Bentler PM. EQS: Structural equations program manual [Computer Program]. Published online 1995.

150. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Struct Equ Model A Multidiscip J*. 1999;6(1):1-55.

151. McLachlan G, Peel D. *Finite Mixture Models*. Wiley; 2004.

152. Titterington DM, Smith AFM, Makov UE. *Statistical Analysis of Finite Mixture Distributions*. Wiley; 1985.

153. Muthén B, Shedden K. Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics*. 1999;55(2):463-469.

154. Berlin KS, Parra GR, Williams NA. An Introduction to Latent Variable Mixture Modeling (Part 2): Longitudinal Latent Class Growth Analysis and Growth

Mixture Models. *J Pediatr Psychol*. 2014;39(2):188-203.

155. Muthén B. The potential of growth mixture modelling. *Infant Child Dev*. 2006;15(6):623-625.

156. Bauer DJ, Curran PJ. The integration of continuous and discrete latent variable models: Potential problems and promising opportunities. *Psychol Methods*. 2004;9(1):3-29.

157. Magidson J, Vermunt JK. Latent class models. In: Kaplan D, ed. *The SAGE Handbook of Quantitative Methodology for the Social Sciences*. Sage; 2004:175-198.

158. Enders CK, Tofighi D. The Impact of Misspecifying Class-Specific Residual Variances in Growth Mixture Models. *Struct Equ Model A Multidiscip J*. 2008;15(1):75-95.

159. Kreuter F, Muthen B. Longitudinal Modeling of Population Heterogeneity: Methodological Challenges to the Analysis of Empirically Derived Criminal Trajectory Profiles. In: Hancock GR, Samuelsen KM, eds. *Advances in Latent Variable Mixture Models*. Information Age Publishing, Inc; 2007:53-75.

160. Frankfurt S, Frazier P, Syed M, Jung KR. Using Group-Based Trajectory and Growth Mixture Modeling to Identify Classes of Change Trajectories. *Couns Psychol*. 2016;44(5):622-660.

161. Hipp JR, Bauer DJ. Local solutions in the estimation of growth mixture models. *Psychol Methods*. 2006;11(1):36-53.

162. Wang M, Bodner TE. Growth Mixture Modeling: Identifying and Predicting Unobserved Subpopulations With Longitudinal Data. *Organ Res Methods*. 2007;10(4):635-656.

163. Akaike H. A new look at the statistical model identification. *IEEE Trans Automat Contr*. 1974;19(6):716-723.

164. Schwarz G. Estimating the Dimension of a Model. *Ann Stat*. 1978;6(2):461-464.

165. Lo Y, Mendell NR, Rubin DB. Testing the number of components in a normal mixture. *Biometrika*. 2001;88(3):767-778.

166. Nylund KL, Asparouhov T, Muthén BO. Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Struct Equ Model*. 2007;14(4):535-569.

167. Tein J-Y, Coxe S, Cham H. Statistical Power to Detect the Correct Number of Classes in Latent Profile Analysis. *Struct Equ Modeling*. 2013;20(4):640-657.

168. Muthén B. Latent Variable Analysis: Growth Mixture Modeling and Related Techniques for Longitudinal Data. In: *Handbook of Quantitative Methodology for the Social Sciences*. Sage; 2004:345-368.

169. Andruff H, Carraro N, Thompson A, Gaudreau P, Louvet B. Latent Class Growth Modelling: A Tutorial. *Tutor Quant Methods Psychol*. 2009;5(1):11-24.

170. Berlin KS, Williams NA, Parra GR. An introduction to latent variable mixture modeling (part 1): overview and cross-sectional latent class and latent profile analyses. *J Pediatr Psychol*. 2014;39(2):174-187.

171. Lubke G, Neale MC. Distinguishing Between Latent Classes and Continuous Factors: Resolution by Maximum Likelihood? *Multivariate Behav Res*. 2006;41(4):499-532.

172. Feldman BJ, Masyn KE, Conger RD. New approaches to studying problem behaviors: a comparison of methods for modeling longitudinal, categorical adolescent drinking data. *Dev Psychol*. 2009;45(3):652-676.

173. van de Schoot R, Sijbrandij M, Winter SD, Depaoli S, Vermunt JK. The GRoLTS-checklist: Guidelines for reporting on latent trajectory studies. *Struct Equ Model*. 2017;24(3):451-467.

174. Lanza ST, Cooper BR. Latent Class Analysis for Developmental Research. *Child Dev Perspect*. 2016;10(1):59-64.

175. Hickendorff M, Edelsbrunner PA, McMullen J, Schneider M, Trezise K. Informative tools for characterizing individual differences in learning: Latent class, latent profile, and latent transition analysis. *Learn Individ Differ*. 2018;66:4-15.

176. Brzinsky-Fay C, Kohler U. New Developments in Sequence Analysis. *Sociol Methods Res*. 2010;38(3):359-364.

177. Mikolai J, Lyons-Amos M. Longitudinal methods for life course research: A comparison of sequence analysis, latent class growth models, and multi-stateevent history models for studying partnership transitions. *Longit Life Course Stud*. 2017;8(2):191-208.

178. Liao TF, Bolano D, Brzinsky-Fay C, et al. Sequence analysis: Its past, present, and future. *Soc Sci Res*. 2022;107:102772.

179. Han Y, Liefbroer A, Elzinga C. Comparing methods of classifying life courses: Sequence analysis and latent class analysis. *Longit Life Course Stud*. 2017;8(4):319-341.

180. Hubert LJ, Levin JR. A general statistical framework for assessing categorical clustering in free itrecall. *Psychol Bull*. 1976;83(6):1072-1080.

181. Nylund-Gibson K, Garber AC, Carter DB, et al. Ten frequently asked questions about latent transition analysis. *Psychol Methods*. 2023;28(2):284-300.

182. Szklo M, Nieto J. *Epidemiology: Beyond the Basics*. Third. Jones & Bartlett Learning, LLC; 2014.

183. NHS Digital. *Fit Notes Issued by GP Practices, England March 2021*.; 2021.

184. Hall SM. *Everyday Life in Austerity: Family, Friends and Intimate Relations*. First. Palgrave Macmillan; 2019.

185. Stuckler D, Reeves A, Loopstra R, Karanikolos M, McKee M. Austerity and health: the impact in the UK and Europe. *Eur J Public Health*.

2017;27(suppl_4):18-21.

186. Knapp M. Mental health in an age of austerity. *BMJ Ment Heal*. 2012;15(3):54-55.

187. House of Commons Library. *Key Issues for the 2015 Parliament*.; 2015.

188. NICE. Low back pain and sciatica in over 16s: assessment and management. NICE guideline [NG59]. Published 2016. Accessed March 13, 2023. https://www.nice.org.uk/guidance/ng59

189. Gabbay M, Shiels C, Hillage J. The issue of "long-term" fit notes for depression in the UK: patient, GP and general practice variation. *J Ment Heal*. 2020;29(6):657-664.

190. Department of Health and Social Care. Online isolation notes launched - providing proof of coronavirus absence from work. Published 2020. Accessed March 13, 2023. https://www.gov.uk/government/news/online-isolation-notes-launched-providing-proof-of-coronavirus-absence-from-work

191. HM Revenue & Customs. Check if you can claim for your employees' wages through the Coronavirus Job Retention Scheme. Published 2020. Accessed March 13, 2023. https://www.gov.uk/guidance/claim-for-wage-costs-through-the-coronavirus-job-retention-scheme

192. Pattyn E, Verhaeghe M, Bracke P. The gender gap in mental health service use. *Soc Psychiatry Psychiatr Epidemiol*. 2015;50(7):1089-1095.

193. McManus S, Bebbington P, Jenkins R, Brugha T. *Mental Health and Wellbeing in England: Adult Psychiatric Morbidity Survey 2014*.; 2016.

194. Chowdhury R, Shah D, Payal AR. Healthy Worker Effect Phenomenon: Revisited with Emphasis on Statistical Methods - A Review. *Indian J Occup Environ Med*. 2017;21(1):2-8.

195. ONS. What are the regional differences in income and productivity? Published 2021. https://www.ons.gov.uk/visualisations/dvc1370/

196. Huggins R, Thompson P, Prokop D. *UK Competitiveness Index 2021*.; 2021.

197. Reeves A, Basu S, McKee M, Marmot M, Stuckler D. Austere or not? UK coalition government budgets and health inequalities. *J R Soc Med*. 2013;106(11):432-436.

198. Lewis M, Wynne-Jones G, Barton P, et al. Should general practitioners issue a sick certificate to employees who consult for low back pain in primary care? *J Occup Rehabil*. 2015;25(3):577-588.

199. Strauss VY, Jones PW, Kadam UT, Jordan KP. Distinct trajectories of multimorbidity in primary care were identified using latent class growth analysis. *J Clin Epidemiol*. 2014;67(10):1163-1171.

200. CPRD. *Small Area Level Data Based on Patient Postcode. Documentation and Data Dictionary (Set 22/January 2022)*.; 2022.

201. Bhatt SP, Kim Y, Harrington KF, et al. Smoking duration alone provides stronger risk estimates of chronic obstructive pulmonary disease than pack-years. *Thorax*. 2018;73(5):414-421.

202. NICE. Obesity: identification, assessment and management. Clinical guideline [CG189]. Published 2014. Accessed December 28, 2023. https://www.nice.org.uk/guidance/cg189/ifp/chapter/obesity-and-being-overweight

203. CPRD. *CPRD Aurum Data Specification. Version 3.1.*; 2023.

204. Bedson J, Belcher J, Martino OI, et al. The effectiveness of national guidance in changing analgesic prescribing in primary care from 2002 to 2009: an observational database study. *Eur J Pain*. 2013;17(3):434-443.

205. Daunt R, Curtin D, O'Mahony D. Polypharmacy stewardship: a novel approach to tackle a major public health crisis. *Lancet Heal Longev*. 2023;4(5):e228-e235.

206. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis*. 1987;40(5):373-383.

207. Charlson ME, Carrozzino D, Guidi J, Patierno C. Charlson Comorbidity Index: A Critical Review of Clinimetric Properties. *Psychother Psychosom*. 2022;91(1):8-35.

208. Vermunt JK. Latent Class Modeling with Covariates: Two Improved Three-Step Approaches. *Polit Anal*. 2010;18(4):450-469.

209. Asparouhov T. Sampling Weights in Latent Variable Modeling. *Struct Equ Model A Multidiscip J*. 2005;12(3):411-434.

210. Wang H, Peng J, Wang B, et al. Inconsistency Between Univariate and Multiple Logistic Regressions. *Shanghai Arch psychiatry*. 2017;29(2):124-128.

211. Lagerveld SE, Brenninkmeijer V, Blonk RWB, Twisk J, Schaufeli WB. Predictive value of work-related self-efficacy change on RTW for employees with common mental disorders. *Occup Environ Med*. 2017;74(5):381-383.

212. Cornelius LR, van der Klink JJL, Groothoff JW, Brouwer S. Prognostic factors of long term disability due to mental disorders: A systematic review. *J Occup Rehabil*. 2011;21(2):259-274.

213. Brouwer S, Amick B, Lee H, Franche R-L, Hogg-Johnson S. The Predictive Validity of the Return-to-Work Self-Efficacy Scale for Return-to-Work Outcomes in Claimants with Musculoskeletal Disorders. *J Occup Rehabil*. 2015;25(4):725-732.

214. Black O, Keegel T, Sim MR, Collie A, Smith P. The effect of self-efficacy on return-to-work outcomes for workers with psychological or upper-body musculoskeletal injuries: A review of the literature. *J Occup Rehabil*. 2018;28(1):16-27.

215. Office for National Statistics (ONS). *INAC01 SA: Economic Inactivity by Reason (Seasonally Adjusted)*.; 2024.

216. DWP, DHSC. Guidance. WorkWell prospectus: guidance for Local System Partnerships. Published 2024. https://www.gov.uk/government/publications/workwell/workwell-prospectus-guidance-for-local-system-partnerships

217. UCL Institute of Health Equity. *Health Inequalities, Lives Cut Short*.; 2024.

218. Grimani A, Aboagye E, Kwak L. The effectiveness of workplace nutrition and physical activity interventions in improving productivity, work performance and workability: a systematic review. *BMC Public Health*. 2019;19(1):1676.

219. NHS England. What are integrated care systems? Accessed February 20, 2024. https://www.england.nhs.uk/integratedcare/what-is-integrated-care/

220. Department of Health and Social Care, Department for Work and Pensions. Employment support launched for over a million people. Published 2023. Accessed February 20, 2024. https://www.gov.uk/government/news/employment-support-launched-for-over-a-million-people

221. CPRD. Introduction to CPRD. Published 2022. Accessed January 11, 2024. https://cprd.com/introduction-cprd

222. Cancelliere C, Donovan J, Stochkendahl MJ, et al. Factors affecting return to work after injury or illness: best evidence synthesis of systematic reviews. *Chiropr Man Therap*. 2016;24:1-23.

223. Office for National Statistics (ONS). Labour Force Survey. Accessed February 21, 2024. https://www.ons.gov.uk/surveys/informationforhouseholdsandindividuals/householdandindividualsurveys/labourforcesurvey

Appendix A

Systematic Review MEDLINE (OVID) Search Strategy

Table A.1. Systematic Review Search Strategy and Results using MEDLINE (OVID); run on 8th April 2021

	Search Term	Results
1	ABSENTEEISM/	9,264
2	(absenteeism).ti,ab	6,084
3	exp Rehabilitation, Vocational/	10,347
4	((job OR work OR occupation* OR vocat*) ADJ2 (absen* OR rehab* OR adj* OR participation OR incapacity OR leave OR return)).ti,ab	19,817
5	Sick Leave/	6,051
6	(sick* ADJ2 (leave OR absen* OR day OR note OR cert* OR pay* OR paid)).ti,ab	8,315
7	(fit* ADJ2 note).ti,ab	67
8	Return to Work/	2,859
9	("return to work" OR rtw OR (work ADJ2 resumption) OR (work ADJ2 re-entry) OR (work ADJ2 return)).ti,ab	9,963
10	((long-term OR "long term" OR longterm) ADJ2 (sick* OR abs*)).ti,ab	2,805
11	((job OR work OR occupation*) ADJ2 (ability OR able OR disab* OR capacity)).ti,ab	14,133

12	(incapacity).ti,ab	3,188
13	1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12	60,436
14	exp Musculoskeletal Diseases/	1,108,197
15	(musculoskeletal ADJ2 (disease* OR pain OR injur* OR disorder OR disorders OR condition*)).ti,ab	22,291
16	exp Pain/	407,149
17	(pain ADJ2 (back OR lumbar OR hand OR knee OR "joint chronic" OR persistent OR "long term" OR long-term OR longterm OR widespread)).ti,ab	73,268
18	(pain ADJ2 low* back).ti,ab	30,966
19	(arthr* OR osteoarthr*).ti,ab	386,704
20	14 or 15 or 16 or 17 or 18 or 19	1,633,423
21	Mental Health/	42,515
22	exp Mental Disorders/	1,277,486
23	(mental ADJ2 (health OR disorder* OR illness*)).ti,ab	209,781
24	exp Anxiety/	90,317
25	(anxiety).ti,ab	204,123
26	Depression/	125,800
27	(depression).ti,ab	349,329
28	(psychiatric ADJ2 illness*).ti,ab	9,511
29	21 or 22 or 23 or 24 or 25 or 26 or 27 or 28	1,693,563

30	cohort studies/ or follow-up studies/ or longitudinal studies/ or prospective studies/	1,445,698
31	(longitudinal or prospective or cohort).ti,ab	1,279,746
32	Observational Study/	95,871
33	(observational).ti,ab	197,238
34	(trajector* or pattern*).ti,ab	1,400,732
35	Latent Class Analysis/	1,096
36	(latent ADJ2 (class or transition)).ti,ab	7,371
37	30 or 31 or 32 or 33 or 34 or 35 or 36	3,491,528
38	20 or 29	3,248,909
39	13 and 37 and 38	6,723

Notes:

1. '.ti,ab' is used to search only in titles and abstracts

2. Subject Headings are searched for wherever '/' is used, this is in accordance with the medical thesaurus: Medical Subject Headings (MeSH)

3. Medical Subject Headings are exploded to include all child Subject Headings that fall in the same branch as the designated Subject Heading by use of the term 'exp'

4. An asterisk ('*') represents truncation. For example, occupation* searches for the terms: occupation, occupations, occupational etc.

4. Using quotation marks only allows the exact term to be searched (as specified in full in the text string)

5. 'ADJ2' allows up to two words to appear between the designated search terms (they can appear in any order)

6. Boolean operators ('and', 'or') are used to combine and restrict the search using mathematical logic

Appendix B

Newcastle-Ottawa Scale (NOS) Tool Original Template

Below are the questions that constitute the original NOS tool, of which an adapted version is later used to assess risk of bias in the systematic review of this thesis.

NEWCASTLE - OTTAWA QUALITY ASSESSMENT SCALE

COHORT STUDIES

Note: A study can be awarded a maximum of one star for each numbered item within the Selection and Outcome categories. A maximum of two stars can be given for Comparability.

Selection

1) Representativeness of the exposed cohort

a) truly representative of the average _____ (describe) in the community *

b) somewhat representative of the average _____ in the community *

c) selected group of users eg nurses, volunteers

d) no description of the derivation of the cohort

2) Selection of the non exposed cohort

a) drawn from the same community as the exposed cohort *

b) drawn from a different source

c) no description of the derivation of the non exposed cohort

3) Ascertainment of exposure

a) secure record (e.g., clinical records) *

b) structured interview *

c) written self report

d) no description

4) Demonstration that outcome of interest was not present at start of study

a) yes *

b) no

Comparability

1) Comparability of cohorts on the basis of the design or analysis

a) study controls for _____ (select the most important factor) *

b) study controls for any additional factor * (This criteria could be modified to indicate specific control for a second important factor.)

Outcome

1) Assessment of outcome

a) independent blind assessment *

b) record linkage *

c) self report

d) no description

2) Was follow-up long enough for outcomes to occur

a) yes (select an adequate follow up period for outcome of interest) *

b) no

3) Adequacy of follow up of cohorts

a) complete follow up - all subjects accounted for *

b) subjects lost to follow up unlikely to introduce bias - small number lost - > ____ %

(select an adequate %) follow up, or description provided of those lost) *

c) follow up rate < ____% (select an adequate %) and no description of those lost

d) no statement

Appendix C

Adapted Risk of Bias Tool

Below are the final set of five questions that constituted the adapted risk of bias tool used in the systematic review of this thesis, based on the Newcastle-Ottawa Scale (NOS). Only one option was permitted to be elected for each question.

A star system was used to rate risk of bias, with studies awarded a maximum of one star for each question, and a total of 5 stars thus being possible (with higher stars indicative of higher quality studies). Risk of bias was then graded into categories, with 0-1 stars indicating a study with low risk of bias, 2-3 stars indicating medium risk of bias, and 4-5 stars indicating low risk of bias.

1) Representativeness of the study cohort (selection bias)

a) truly representative *

b) somewhat representative *

c) selected group of users (e.g., nurses, volunteers)

d) no description of the derivation of the study cohort

2) Ascertainment of work absence (measurement bias)

a) electronic health records *

b) structured interview *

c) written self-report

d) no description

3) Assessment of outcome (outcome bias)

a) independent blind assessment *

b) record linkage using electronic health records *

c) self-report

d) no description

4) Was follow-up long enough for outcomes to occur? (outcome bias)

a) yes *

b) no

5) Adequacy of follow up of cohorts (outcome bias)

a) complete follow up - all subjects accounted for *

b) subjects lost to follow up unlikely to introduce bias (small number lost - <=10 %, or description provided of those lost, clearly demonstrating that the group of patients lost and those retained were similar in important participant characteristics) *

c) follow up rate < 90%, or no description of those lost (or description provided, but group of patients lost appear dissimilar to group retained)

d) no statement

Appendix D

Further Systematic Review Results

Table D.1. More Detailed Summary of Characteristics of Included Studies in Systematic Review

First Author, Publication Year	Country	Missing Data Description	Inclusion Criteria	Exclusion Criteria	Baseline Participant Characteristics	Work Absence Reason Definition	Work Absence Database Used
Pedersen, 2016	Denmark	n=300 participants were excluded, pre-analysis, due to missing self-reported data on: - sickness absence reason, n=20 - education, n=31 - employment, n=123 - return to work expectations, n=126 A further n=452 participants also excluded pre-analysis, due to no record of sick leave being found in the DREAM database, thus in total n=752 missing	- On sick leave for 4-8 weeks at the time of baseline questionnaire issuance (September 2012 - January 2014) - 18 to 64 years of age - SCL-8 AD score ≥5	- Individuals who did not communicate in Danish - On sick leave due to mental health for > 3 consecutive months during the year preceding first point of contact (September 2012 - January 2014) - Pregnant individuals - Had a supported	- Men, 33.7% - Mean age, 42.3 years - Employed, 76.2% - Reason for mental health sickness absence diagnosis: stress and burnout, 71.2%; depression, 55.9%; anxiety, 30.1%; other mental illness, 11.0% (note: total is not 100% as multiple simultaneous sickness absence reasons were possible).	- Self reported mental health states of: anxiety, depression, stress and burnout, or other mental illness	- The Danish National Labour Market Authority's DREAM database for economic information relating to sickness absences, although used self-reported questionnaire data to determine the corresponding reason for

		data exclusions. Furthermore, n=4 participants lost to follow-up due to death or emigration.		job/was in job training/in rehabilitation - Retired individuals			sickness absence, which may be less accurate than using objective data
Farrants, 2018	Sweden	None stated	- Age 16-64 years - Living in Sweden at 31st December 2009 - New sick leave absence due to depression in first 6 months of 2010 - Not been on disability pension, nor reached maximum number of sick leave days	- Previous diagnosis of depressive episode (measured using ICD-10 code F33)	- Men, 31.6% - Median age, 42 years - In paid work, 84.1% - Married/cohabiting, 48.1% - Majority born in Sweden, 84.5% - Born outside of EU, 10.1% - Education >= elementary level, 86.8%	- Depressive episode, measured using ICD-10 code F32	- A database from the National Social Insurance Agency of Sweden: MicroData for Analysis of the Social Insurance database (MiDAS)

			(365-914 days) in the year before sickness absence - Survived first 180 days of the sick leave absence - Sick leave absence of >= 21 days				
McLeod, 2018	Canada	n=462 work-related MSD lost time claims excluded, due to: - missing firm size, n=67 - missing wage, n=10 - missing gender, n=5 - combination of missing data, n=380	- An accepted disability claim due to work-related musculoskeletal disorder (MSD), during the time period 1 January 2010 to 31st December 2012 - First MSD lost time claim only (during the time period 1 January 2010 to 31st December 2012); i.e. further sickness	- Age < 15 or > 65 years - Multiple jobholders - Incomplete claims - Claims related to previous non-MSD claims - Claims from self-insured industry sectors - Claims related to fatal injuries - Claims with missing data	- Men, 63% - Mean age, 41 years - Mean annual wage prior to injury, approximately CAD$40,000 - 65% of sample had history of prior claims - More than a third of all occupations were in trades, and the most common industry sectory was service (42%) - Median firm size 144 FTE workers - Sprains and strains were most common type of MSD (79%),	- Work related musculoskeletal disorders, defined using ICD-9-Clinical Modification codes and National Work Injury Statistics Program (NWISP) WorkSafeBC Nature of Injury codes	- A Canadian health maintenance organi sation (HMO) workers' compensation scheme, known as WorkSafeBC, which is funded by insurance premiums paid by employers

			absences post a successful full return-to-work in this period were excluded - At least one full day off work		followed by fractures (10%), and then musculoskeletal diseases (10%), and with dislocations as the least common MSD type (2%)		
Farrants, 2019	Sweden	None stated	- Age 16-64 years - Living in Sweden at 31st December 2009 - New sick leave absence due to osteoarthritis in first 6 months of 2010 - Not been on disability pension, nor reached maximum number of sick	- No further exclusions stated (in addition to those contradicting inclusion criteria)	- Men, 47.6% - Mean/median age not reported, but over 60% of cohort >= 55 years old - In paid work, 94.2% - Married/in registered partnership, 57.7% - Majority born in Sweden, 90.0% - Education >= elementary level, 79.9%	- Osteoarthritis, measured using ICD-10 code M15-19	- A database from the National Social Insurance Agency of Sweden: MicroData for Analysis of the Social Insurance database (MiDAS)

	leave days (365-914 days) in the year before sickness absence
	- Survived first 180 days of the sick leave absence
	- Sick leave absence of >= 21 days

Spronken, 2020	The Netherlands	Mentioned in the supplementary materials that n=1833 sickness absence records containing errors were deleted, not clear how many of these were due to missing data specifically	- An episode of sickness absence due to mental health in 2014 (and only the first one was selected, in the case of multiple instances in 2014) - Sickness absence duration from 29 to 730 days	- Employees with more than one employer/contract - Employees whose contract ended within 7 days after the end of the sickness absence period	- Men, 47.3% - Mean age, 41.8 years - Adjustment disorder was the most common mental health problem, 57.6% prevalence - Majority of employees worked in profit sector, 75.9% - Smaller organisations of <=50 employees had a prevalence of 52.8%	- Mental health problems, defined using ICD-10. Inclusions were: - Employees with stress complaints (R45), - Emotional sleeping disorders (F51.9), - Somatoform disorders (F45.0, F45.4, F45.9), - *Adjustment disorders (F43.2, Z73.0), - Reactions to severe stress (F43.1, F43.9), - Anxiety disorders (F41.0, F41.1, F40.0, F40.1, F41.9), - Personality disorders (F60.0, F60.1, F60.2, F60.3, F60.4, F60.6, F60.7, F60.8, F60.9), - Mood disorders (F30.9, F31.9, F32.9, F34.1, F39), - Addictions due to psychoactive substances (F10.9, F11.9, F15.9, F19.9), - Organic psychoses (F09), - Non-organic psychoses (F20.9, F25.9, F29), - and Other mental disorders (F48.0, F48.8, F42.9, F44.9, F50.9, F53.9, F63.0, F79, F99) *Note: Burnout (Z73.0) belongs to the category 'adjustment disorders'.	- The largest Dutch occupational health service, HumanTotalCare

Rysstad, 2023	Norway	Small amount of missing data for prognostic factors (<2%), and 'little difference between responders and non-responders'. Complete case analysis performed.	- All Norwegian workers on sick leave for at least 4 weeks due to a musculoskeletal disorder - Working age (18-67 years)	- On sick leave for less than 4 weeks - Unemployed - Insufficient Norwegian or English skills	- Median age 50.1 years (range 18.6-67.9) - 56.3% women	- Musculoskeletal disorders, as defined by a diagnosis within the musculoskeletal (L) chapter of the International Classification of Primary Care, second edition (ICPC-2)	- The National Sick Leave Registry (a Norwegian sick leave database)

Table D.2. Full Summary of Derived Work Absence Trajectories from Studies Included in Systematic Review (Including Descriptive Participant Characteristics within Trajectories)

First Author, Publication Year	Trajectory Number	Trajectory Name and Description	Trajectory Prevalence, %	Trajectory Participant Characteristics
Pedersen, 2016	1	Sickness absence (almost 100% of individuals on sickness absence for approximately the first half of follow-up, before slowly reducing to around 50% by end of follow-up)	44.0	No trajectory participant characteristics reported (for the group absent at baseline due to a MH condition)
	2	Fast RTW (initial 100% sickness absence state decreased rapidly to 0% by around month 3, which was then sustained until end of follow-up. Individuals who exited the sickness absence state largely entered a work state).	21.9	
	3	Slow RTW (similar to fast RTW trajectory, except that 100% sickness absence state was sustained for approximately first 3 months, before decreasing steadily to 0% by around month 7, which was sustained until end of follow-up)	14.4	

4	Sickness absence/temporary support (after around 4 months with approximately all individuals on a sickness absence, rapid decrease to 0% by around month 8, which was sustained until end of follow-up. Participants who exited the sickness absence state largely entered a temporary support state)	5.4
5	Temporary support (similar to trajectory class 4, except no sustained 100% sickness absence state in the initial months, instead there was an immediate rapid decrease to 0% of individuals on a sickness absence, which was reached by around month 4).	5.1
6	Unemployment (moderately fast decrease to 0% sickness absence by around month 5, which then remained low until end of follow-up. Individuals exited sickness absence largely to enter an unemployment state).	4.4
7	Permanent support (very similar to the unemployment trajectory, except individuals exiting sickness absence largely entered a permanent support state).	2.4
8	Relapse (% of individuals on sickness absence decreased from 100% to around 20% after first 3 months, before rapidly increasing back to 100% by around month 8, and then slowly decreasing again)	2.4

	1	Decrease to 0 (monthly days of SA/DP) after 4 months	43.0	- This trajectory had the highest proportion of people with no sick leave prior to the index spell (76.6%), and one of the lowest proportions of specialist in- and out-patient care (85.6% had no outpatient specialist visit in the first 21 days, and 94.5% had no inpatient care during the first 21 days) - One of the highest proportions of younger individuals (<35 years); 30.6%
	2	Decrease to 0 after 9 months	22.0	- One of the highest proportions of younger individuals (<35 years); 32.6%
Farrants, 2018	3	Constant high (at around 30 days of SA/DP per month)	11.0	- One of the highest proportions of individuals on full-time sick leave at the start of the sickness absence; 96.5%. Due to this, the cluster started with a high average of 30 days of work disability per follow-up month, which continued and remained fixed until around month 9, after which it decreased slightly. - High proportion of men (35.4%) - High proportion of individuals born outside of the Nordic countries (18.9%) - High proportion of individuals with only elementary level education (18.6%) - High proportion of unemployed individuals (33.5%) - High proportion of healthcare consumption, e.g. a high proportion of at least one day of inpatient care, before index sickness absence (16.8%), and during first 21 days (11.4%) - High proportion of individuals that received inpatient care due to a suicide attempt, before index sickness absence (1.7%), and during first 21 days (1.2%) - One of the highest proportions of individuals that had a previous depression-related sick leave (11.5%), yet also one of the highest

			proportions of individuals that hadn't purchased anti-depressants prior to the sickness absence (57.7%)
4	Decrease, then high increase (decrease from a high net days of SA/DP of around 22 per month at the start of follow-up, to around 14 by month 5, before increasing again)	9.0	- One of the highest proportions of individuals with prior sick leave (39.7%) - One of the highest proportions of individuals that had a previous depression-related sick leave (13.8%), yet also one of the highest proportions of individuals that hadn't purchased anti-depressants prior to the sickness absence (57.3%) - Highest proportion of individuals with part-time sick leave at the start of the sickness absence (27.2%) - High proportion of specialist in- and out-patient care due to somatic diagnosis (e.g. 46.1% had a specialised outpatient visit prior to the sickness absence due to somatic diagnosis, and 11.0% had a corresponding inpatient visit)

	5	Slow decrease (started off with high net SA/DP days of 30 per month, but decreased steadily and continuously to around 3 days per month by end of follow-up)	9.0	- Highest proportion of individuals born in Sweden (87.2%) - One of the highest proportions of individuals on full-time sick leave at the start of the sickness absence (96.5%). Due to this, the cluster started with a high average of 30 days of work disability for the first two months of follow-up, after which it decreased sharply over the coming months. - A high proportion of outpatient care due to somatic diagnoses; 45.7%
	6	Decrease, then low increase (decreased from around 17 net days of SA/DP per month at the start, to around 1 day per month by month 4, which remained until around month 6, before a steady increase was observed until end of follow-up)	6.0	- One of the highest proportions of specialist in- and out-patient care (83.1% had no outpatient specialist visit in the first 21 days, and 94.6% had no inpatient care during the first 21 days)
McLeod, 2018	1	Early-sustained RTW (reached a sustained state of RTW by the 1st month)	49.7	- Higher percentage of younger individuals - Lower percentage of individuals with upper or lower extremity fractures, torso fractures, or dislocations. - Higher percentage with lower extremity sprains and strains, or back sprains and strains.
	2	Short-delayed RTW (reached a sustained state of RTW during months 2–6)	30.6	- Higher percentage of individuals with upper or lower extremities fractures - Higher percentage of older individuals
	3	Early NRTW (reached end state of NRTW within the first 6 months)	6.7	- Higher percentage of younger individuals - Higher percentage earning the least amount (<$20.000)
	4	Long-delayed RTW preceded by SA (reached a sustained state of RTW by months 7–13, with preceding events predominantly a SA state)	4.2	

	5	Late NRTW (reached end state of NRTW by months 7–13)	3.1	
	6	Constant SA (remained in SA state throughout follow-up)	3.0	- Considerably higher percentage of individuals with torso fractures
	7	Deferred SA (reached a sustained state of SA anytime during months 2–13)	0.9	
	8	Long-delayed RTW preceded by MRTW (reached a sustained state of RTW by months 7–13, with preceding events predominantly a MRTW state)	0.8	
	9	Unclassifiable	1.1	
Farrants, 2019	1	Fast decrease (had no/very little SA/DP days per month after 4 months of follow-up)	36.0	- This cluster had the highest proportion of younger people (<= 49 years); 27.6% - Highest proportion of individuals with a university education (31.9%) - Highest proportion of individuals in employment (i.e. either employed or self-employed); 95.8% - Highest proportion of specialised outpatient healthcare during the first 21 days of sickness absence (33.6%), and of osteoarthritis-related specialised outpatient healthcare (23.7%) - One of the lowest proportions of previous sickness absence (75.5% with 0 net days of previous sickness absence)

2	Medium fast decrease (had no SA/DP days per month after 5 months of follow-up)	29.0	- One of the lowest proportions of previous sickness absence (76.3% with 0 net days of previous sickness absence) - Largest proportion of individuals born in Sweden, 92.5% - Almost the whole trajectory had full-time SA at the start of the index spell (99.7%) - Highest proportion of inpatient healthcare during the previous year (25.1%), and previous inpatient healthcare due to osteoarthritis (21.2%) - Highest proportion of individuals who purchased prescribed analgesics during the first 21 days (84.8%)
3	Slow decrease (had no SA/DP days per month after 10 months of follow-up)	15.0	- Smallest proportion of specialised outpatient healthcare (24.8%), and osteoarthritis related specialised outpatient healthcare in the first 21 days (17.7%)
4	Fluctuating (started off with around 20 SA/DP days in the first month, which then decreased over the first few months, before steadily increasing again from month 5)	12.0	- One of the smallest proportions of younger individuals (aged 20-39 years); (3.3%) - One of the largest proportions of individuals born outside the EU25 (5.3%) - Largest proportion of self-employed individuals (8.9%) - Largest proportion of part-time SA at the beginning of sickness absence (30.5%, twice as high as the second largest, slow decrease, 14.5%).

	5	Late decrease (started off with high SA/DP days per month of >25, which was sustained for the first 9 months, before steadily decreasing to about 15 days per month at end of follow-up)	8.0	- One of the smallest proportions of younger individuals (aged 20-39 years); (2.6%) - One of the largest proportions of individuals born outside the EU25 (6.1%) - Highest proportion of individuals with previous SA due to mental diagnoses (4.2%) - Highest proportion of individuals living in rural areas (40.1%) - Highest proportion of unemployed individuals (12.9%) - Highest proportion of individuals who had any previous SA (42.5%), - Highest proportion of individuals who had previously been hospitalised due to mental diagnoses (1.1%), and who had previously purchased prescribed analgesics (60.2%) - But smallest proportion of individuals who had purchased prescribed analgesics during the first 21 days (49.3%).
Spronken, 2020	1	Fast RTW with little chance of relapse (average of 136 days of sickness absence follow-up and 1.96 transitions before full RTW of 100% of contract hours achieved)	49.5	- Younger age (mean=41.9 years) than slower RTW trajectories - Higher proportion of males (48%) than slower RTW trajectories - Higher proportion of stress complaints (18%) than slower RTW trajectories - Highest proportion of adjustment orders (62%) - Lower proportion of burnout (8%) than slower RTW trajectories - Lower proportion of depression and mood disorders (7%, and 8%, respectively) than slower RTW trajectories

2	Slow RTW with little chance of relapse (average of 402 days and 2.47 transitions before full RTW achieved)	20.8	- Oldest age (mean=43.4 years) than other trajectories - Lower proportion of males (45%) than faster RTW trajectories - Lowest proportion of stress complaints (5%) - Higher proportion of adjustment orders (53%) than the fastest, but not the faster RTW trajectories - Highest proportion of burnout (17%) - Higher proportion of depression and mood disorders (21%, and 23%, respectively) than faster RTW trajectories
3	Fast RTW with considerable chance of relapse (average of 194 days and 3.07 transitions before full RTW achieved)	11.1	- Youngest age (mean=39.8 years) compared to other trajectories - Higher proportion of males (48%) than slower RTW trajectories - Higher proportion of stress complaints (19%) than slower RTW trajectories - Higher proportion of adjustment orders (59%) than the fastest and slower RTW trajectories - Lower proportion of burnout (6%) than slower RTW trajectories - Lower proportion of depression and mood disorders (7%, and 8%, respectively) than slower RTW trajectories
4	Slow RTW with considerable chance of relapse (average of 419 days and 3.54 transitions before full RTW achieved)	9.5	- Older age (mean=42.0 years) than faster RTW trajectories - Lowest proportion of males (40%) - Lower proportion of stress complaints (8%) than faster RTW trajectories - Higher proportion of adjustment orders (54%) than the fastest, but not the faster RTW trajectories - Higher proportion of burnout (12%) than faster RTW trajectories - Higher proportion of depression and mood disorders (19%, and 21%, respectively) than faster RTW trajectories

	5	Very fast RTW with very small chance of relapse (average of 49 days and 1.00 transitions before full RTW achieved)	9.1	- Younger age (mean=40.7 years) than slower RTW trajectories - Highest proportion of males (54%) - Highest proportion of stress complaints (37%) - Lowest proportion of adjustment orders (48%) - Lowest proportion of burnout (1%) - Lower proportion of depression and mood disorders (6%, and 6%, respectively) than slower RTW trajectories
Rysstad, 2023	1	Fast decrease (rapid decrease to 0 sickness absence days 4 months from first assessment, then sustained RTW)	27.0	- Lowest median sickness absence days in previous year (30.0 days)
	2	Moderate decrease (slower decrease to approximately 0 sickness absence days by around month 8, then sustained RTW)	22.4	
	3	Persistent high (stable and high number of sickness absence days throughout follow-up)	18.2	- Highest median age of 52.3 years - Highest median sickness absence days in previous year (80.4 days) - Highest median sickness absence days during the one year follow-up (221.1 days) - Highest percentage wanting a new job after the sick leave (36.4%)
	4	Persistent moderate (stable and moderate number of sickness absence days throughout follow-up)	12.8	- High proportion of women (74.3%) - High median sickness absence days during the one year follow-up (111.4 days)
	5	Slow decrease (steady decrease to 0 sickness absence days at around month 11)	12.4	- Lowest median age of 47.1 years - High median sickness absence days during the one year follow-up (120.3 days)
	6	U-shape (fast decrease in sickness absence days in first 4 months, followed by recurrence of absence from month 8 onwards)	7.3	- High proportion of women (75.0%)

Abbreviations: MH = Mental Health; RTW = return to work; NRTW = non return to work; MRTW = modified return to work; SA = sickness absence; DP = disability pension.

Appendix E

Original data request protocol, submitted to CPRD via eRAP online system on

17/11/2021, and approved on 22/12/2021

General information

Study title
Longitudinal trajectories of work absence in patients with musculoskeletal and, or, mental health conditions

Research area
Health Services Delivery, Methodological

Does this protocol describe an observational study using purely CPRD data?
Yes

Additional information from GPs and contact with patients

Does this protocol involve requesting any additional information from GPs, or contact with patients?
No

Research team

Applicant's role
Role: Corresponding applicant
Email: a.s.legha@keele.ac.uk
Name: Mr Amardeep Legha
Statistical experience: Yes
Experience of handling large datasets: No
Experience of practicing in UK primary care: No
Will the applicant be analysing the data? Yes

Chief investigator
Chief investigator's email: g.wynne-jones@keele.ac.uk
Will this person be analysing the data?: No
Status: Confirmed
Name: Dr Gwenllian Wynne-Jones
Statistical experience: No
Experience of handling large datasets: No
Experience of practicing in UK primary care: No

Collaborators
Collaborator's email: k.p.jordan@keele.ac.uk
Will this person be analysing the data?: Yes

Status: Confirmed
Name: Professor Kelvin Jordan
Statistical experience: Yes
Experience of handling large datasets: Yes
Experience of practicing in UK primary care: No

Collaborator's email: v.welsh@keele.ac.uk
Will this person be analysing the data?: No
Status: Confirmed
Name: Dr Victoria Welsh
Statistical experience: No
Experience of handling large datasets: No
Experience of practicing in UK primary care: Yes

Collaborator's email: j.bailey4@keele.ac.uk
Will this person be analysing the data?: Yes
Status: Confirmed
Name: Mr James Bailey
Statistical experience: No
Experience of handling large datasets: Yes
Experience of practicing in UK primary care: No

Collaborator's email: c.m.holdsworth@keele.ac.uk
Will this person be analysing the data?: No
Status: Confirmed
Name: Professor Clare Holdsworth
Statistical experience: No
Experience of handling large datasets: No
Experience of practicing in UK primary care: No

Access to data

Sponsor
Sponsor: Keele University
(Sponsor information is retrieved automatically as the chief investigator's affiliation)

Funding source for the study
Is the funding source for the study the same as Chief Investigator's affiliation? No

Funding source for the study
Funding source for the study: Economic and Social Research Council (ESRC)

Institution conducting the research
Is the institution conducting the research the same as Chief Investigator's affiliation? Yes

Method to access the data
Indicate the method that will be used to access the data: Institutional multi-study licence

Is the institution the same as Chief Investigator's affiliation? Yes

Extraction by CPRD
Will the dataset be extracted by CPRD? No

Multiple data delivery
This study requires multiple data extractions over its lifespan {Empty}

Data processors
Data processor is: Same as the chief investigator's affiliation
Processing: Yes
Accessing: Yes
Storing: Yes
Processing area: UK

Information on data

Primary care data: CPRD Aurum

Do you require data linkages? No - I do not require data linkages

Area level data
Do you require area level data? Yes
Practice level (UK): {Empty}
Patient level (England only): Patient Level Index of Multiple Deprivation

Linkage to a dataset not listed
Are you requesting a linkage to a dataset not listed? No

Patient data privacy
Does any person named in this application already have access to any of these data in a patient identifiable form, or associated with an identifiable patient index? No

Protocol information

Lay summary
Aches and pains, as well as mental health conditions, are one of the biggest causes of work absence. Most people return-to-work reasonably quickly after an episode of healthcare, but around 10 in 100 go on to have a longer-term absence of more than 12 months.

If a sickness absence lasts more than 7 days and sick pay is required, a General Practitioner (GP) can issue a fit note, that contains their recommendations about a potential return to work. Fit note information is recorded in primary care electronic health records. Having access to CPRD allows us to access fit note data, and use it as a measure of work absence, to investigate work absence patterns over time. This is important, because during a consultation with a GP it is often difficult to tell who is at risk of longer-term absence. So we plan to use fit note data in patients with pain or a

mental health condition to see if we can find common patterns of work absence such as having a long absence or returning to work quickly.

We also want to see if the health and sociodemographic characteristics of a person affect their chances of following a particular work absence pattern. For example, are people living in more disadvantaged neighbourhoods more likely to have a longer-term work absence?

Ultimately, the goal is to allow GPs to give more specific support to their patients at first consultation, to aid the return-to-work process.

Technical summary

BACKGROUND

Ability to work is one of the biggest drivers of social inequalities, leading to adverse health and social outcomes. Absence from work due to musculoskeletal and/or mental health conditions accounts for the majority of healthcare costs and productivity losses. Most people return-to-work relatively quickly following an episode of healthcare, but approximately 10% go on to have a longer-term work absence of > 12 months.

Fit notes are statements issued by GPs that record their medical recommendations regarding a potential return-to-work for patients absent for more than 7 days. They are recorded in primary care electronic health records; access to such data potentially allows uncovering of patterns of work absence over time (trajectories). Knowledge of these trajectories and associated characteristics could help GPs better distinguish patients at higher risk of sustained long-term work absence during initial consultation, and thus potentially offer earlier and more targeted intervention to such patients.

AIMS AND OBJECTIVES

1) To derive, and compare using different statistical methods, common longitudinal trajectories of work absence as measured by receipt of fit notes, for a population consulting their GP with a musculoskeletal and/or mental health condition
2) To derive health and sociodemographic characteristics associated with these trajectories

METHODS

For a population absent from work due to musculoskeletal and/or mental health conditions:

Study 1: Derivation of rates and duration of work absence (2010-2021), with differences examined by: sociodemographic characteristics (age, sex, and geographic region).

Study 2: Derivation of trajectories of work absence (2016-2018); contrasted using simple methods of modelling trajectories, against more complex approaches (such as different types of latent class analysis).

Study 3: Multivariable multinomial regression analyses to test association of each derived trajectory with the sociodemographic characteristics specified in study 1, as well as deprivation status, health characteristics, comorbidity, and treatment received.

Outcomes to be measured

STUDY 1
Rates and duration of work absence

STUDY 2
Trajectories of work absence (derivation)

STUDY 3
Trajectories of work absence (association of characteristics)

Objectives, specific aims and rationale

GENERAL OBJECTIVE
To describe the common longitudinal trajectories of work absence as measured by receipt of fit notes, in a population consulting their general practitioner with a musculoskeletal or mental health condition, and to derive profiles of patients within each of these trajectories. This information will be used to help inform a more timely and targeted intervention approach for GPs during first consultation with their patients, potentially making treatments more likely to succeed, and reducing the adverse health and social potential of worklessness.

SPECIFIC AIMS
Undertake the following in patients with a musculoskeletal and/or mental health condition in CPRD Aurum in England:

1) To derive, and compare using different statistical methods, common longitudinal trajectories of work absence. This will test the hypotheses around the existence, and possible identification in primary care, of different trajectories of work absence for patients with a musculoskeletal and/or mental health condition.
2) To derive health and sociodemographic characteristics associated with these trajectories. This will test the hypothesis that patients with certain health and sociodemographic characteristics, comorbidities, and receiving particular treatments, are more likely to belong to certain trajectories of work absence over others.

RATIONALE FOR STUDY
In the UK, it is known that long-term sickness absence is one of the main drivers of social inequalities, which lead to adverse social and economic outcomes for the individual, the employer, and wider society (1). In financial terms, for example, an estimated £16.2 billion was lost in the UK in the year 2018/19 due to work-related injury and ill health (2). Musculoskeletal and mental health conditions account for the majority of sickness absences in the UK, with for example, 27% and 55% of total working days lost due to ill health in 2019/20, due to musculoskeletal and mental health conditions respectively (2). Although research shows that the majority of people do return to work within a short time frame following a medical problem, around 10% are expected to go on to experience longer term absences of > 12 months (3). The lengthier the work absence, the harder it is for the individual to return to work, and the more

adverse health and social problems they experience (4). Work is generally good for physical and mental health, and worklessness is associated with poorer health and well-being (4).

If sickness absence persists for more than seven continuous days, a fit note can be issued. This is a written statement from a GP or other qualified medical practitioner that records the medical advice a patient has received regarding their fitness to work (5). Primary care electronic health records (EHRs), provide a valuable opportunity to examine work absence, by assessing patterns in the issuance of fit notes, through use of trajectory analysis. A trajectory describes the evolution of a repeated measure over time (for example, the course of work absence). This is achieved by detecting subgroups of individuals with similar patterns in a set of longitudinal heterogeneous data. Classification into subgroups is performed such that individuals share more similarities within their subgroup, than outside of the subgroup (6). During an initial GP consultation for sickness absence, it is challenging to determine which patients are at the highest risk of sustained long-term work absence; determining trajectories of work absence and the patient characteristics associated with them may assist with this. However, few studies of work absence trajectories have been conducted, especially for work absence due to musculoskeletal and/or mental health conditions (the two most common causes of work absence). Much of the work absence literature has been focused on investigating risk factors for experiencing a sickness absence; the few studies that do consider participants already on a sickness absence at baseline tend to consider only dichotomous outcomes, such as still having a sickness absence/no return-to-work after: 3 months (7), 12 months (8), or 2 years (9). Dichotomous outcomes may under- or over-estimate work absence (10); instead, using trajectories of work absence utilises the heterogeneity in both speed and duration of return-to-work spells, to form different trajectory clusters, and considers returning to work as a process as oppose to a fixed outcome.

To the best of our knowledge, our study will be the first to explore trajectories of work absence due to musculoskeletal and/or mental health conditions in an English population, as well as the profiles of patients associated with such trajectories. This is important because use of such trajectories can help GPs better understand, at initial consultation, which patients are more likely to undergo a detrimental course of work absence over time, and thus to act on this information by providing more timely and targeted support for such patients.

Study background

Long-term sickness absence in the UK is one of the main drivers of social inequalities, leading to adverse social and health outcomes (1). The costs of long-term sickness absence are significant, for example, an estimated £16.2 billion was lost in the UK in the year 2018/19 due to work-related injury and ill health (2).

Social inequality is the degree to which differences exist amongst groups in society, or "the condition where people have unequal access to valued resources, services, and positions in the society" (11)(p.11). Much research has been conducted on social inequalities that influence health, especially since the 1980 publication of The Black Report (otherwise known as 'the Report of the Working Group on Inequalities in Health') (12). The authors demonstrated that since the inception of the National Health

Service (NHS) in 1947, differences in risk of mortality had not reduced between different social classes, and these health inequalities had actually increased from the 1950s to the 1970s (13). The foremost finding from this report was that this was due to social inequalities rather than the NHS per se, due to differences that for example relate to: income, nutrition, education, and housing (13).

Other reports followed, such as the Whitehall II study which showed that social class (defined as employment grade) was inversely associated with risk of morbidity under a wide range of diseases, and with risk of mortality (14). The ground-breaking Strategic Review of Health Inequalities in England post-2010 (SRHIE, 2010) report also followed, and aimed to devise effective strategies to reduce health inequalities present in England (15). One of the foremost policy recommendations of the SRHIE (2010) concerns employment. Marmot et al set as a specific policy objective, the need to improve access to and quality of good employment, across all social strata. Additionally, they emphasized the need to reduce long-term unemployment across all social strata. This is important because it is known that work is generally good for physical and mental health, and worklessness is associated with poorer health and well-being (4).

Patterns of employment and work absence over time both reflect and reinforce the social gradient, and demonstrate the inequalities of access to labour market opportunities. People who are in lower socioeconomic positions are at a greater risk of unemployment (16), and being unemployed is associated with a greater rate of long-term illness (17), as well as mental illness (18). For the majority of the healthy working population the spiral towards worklessness tends to start with the onset of ill health, if this progresses to a point in which sick pay is required through the state, a fit note will usually need to be obtained from a medical practitioner, usually a GP. The premise behind a fit note is that the right kind of work is generally good for a patient's physical and mental well-being (4), and that it is not necessary to be completely fit to work in many instances (19); hence fit notes are administered with the ultimate goal of helping patients return to work as soon as they can and aiding their recovery (20)(21).

Although research shows that the majority of people do return to work within a short time frame following a medical problem, around 10% are expected to go on to experience longer term absences of > 12 months (3). The lengthier the work absence, the harder it is for the individual to return to work, and the more adverse health and social problems they experience (4). Poor health, and in particular chronic conditions, significantly affect one's ability to work. Chief among these are musculoskeletal and mental health conditions. For example, it was estimated that in the 2016/17 financial year £10.6 billion was spent on mental health related sickness absence in the UK (22). Additionally, it was shown that musculoskeletal disorders accounted for 27% of total working days lost due to ill health in 2019/20, and that 8.9 million working days were lost due to this condition in this time period (23).

Fit notes are recorded in primary care electronic health records. Having access to such data provides a unique opportunity to examine work absence, by assessing patterns in the issuance of fit notes over time (i.e., trajectories). During an initial GP consultation for sickness absence, it is challenging to determine which patients are at the highest risk

of sustained long-term work absence; determining trajectories of work absence and the patient characteristics associated with them may assist with this.

Trajectories of work absence are scarcely studied. To the best of our knowledge, our study will be the first to explore trajectories of work absence due to musculoskeletal and/or mental health conditions in an English population, as well as the profiles of patients associated with such trajectories. This is important because a more complete understanding of the intricate courses of work absence over time for such individuals is hypothesised to be useful in tackling the issue of work absence. Hence, use of trajectories in this context can help GPs to better understand, at initial consultation, which patients are more likely to undergo a detrimental course of work absence over time, and thus to act on this information by providing more timely and targeted support for such patients.

Study type
Study aim 1 will be addressed through a descriptive study.

Study aim 2 will be addressed through a methodological/hypothesis generating study.

Study aim 3 will be addressed through a hypothesis generating study.

Study design
The research will involve three inter-related studies, each mapped to its own objective.

STUDY 1: will involve cross-sectional analyses to establish incidence rate and duration of work absence due to musculoskeletal and/or mental health conditions, and with trends in rates of absence over time to be compared, and differences explored by: sociodemographic characteristics (age, sex, and geographic region).

STUDY 2: will involve a retrospective cohort study whereby trajectories of work absence due to musculoskeletal and/or mental health conditions will be derived; there will also be a methodological component to this study as simple methods of modelling trajectories (such as cumulative duration of work absence), will be contrasted against more complex approaches (such as different types of latent class analysis, including latent class growth analysis, growth mixture modelling, and latent transition analysis).

STUDY 3: will involve a retrospective cohort study using multivariable multinomial regression analyses to test association of each derived trajectory in study 2, with: sociodemographic characteristics (age, sex, geographic region, and deprivation status), health characteristics (body mass index, smoking, and prior and current consultation patterns for musculoskeletal and mental health conditions), comorbidity (defined by Charlson Index, polypharmacy, consultation count, and specific comorbidities), and treatment received (defined as analgesia, anti-inflammatory medication, and anti-depressants).

Feasibility counts
In order to inform feasibility counts of the number of patients with a new fit note in a given time period, a list of Medical Code IDs was first devised using Read/SNOMED codes (provided in the Appendix).

Using this list of Read/SNOMED codes, and applying an estimate of the percentage of total fit notes that were issued due to musculoskeletal conditions (using data from a March 2021 NHS Digital Report) (24), feasibility counts (for patients with minimum age >= 16 years, and at least 2 years prior registration) suggest around 60,000 patients per year with a first fit note due to musculoskeletal conditions (2010-2014), up to around 120,000 per year (2019).

Similarly, feasibility counts suggest around 140,000 patients per year with a first fit note due to mental health conditions (2010-2014), up to around 300,000 per year (2019).

Sample size considerations

Based on the large numbers in the feasibility counts seen in the previous section, for our main study (study 2), 50,000 patients will be randomly selected from the population of patients with a first fit note due to musculoskeletal, or mental health conditions, respectively, for annual cohorts selected from 2016-2018 (further details in the Data/Statistical Analysis section), with each cohort followed-up for 3 years. Hence for each cohort, trajectories of work absence will be calculated based on a total sample of 150,000 patients. Thus, using a normal approximation to the binomial calculation, with this sample size, 95% confidence intervals assuming:

• 10% trajectory prevalence will be (9.85%, 10.15%)
• 30% trajectory prevalence will be (29.77%, 30.23%)
• 50% trajectory prevalence will be (49.75%, 50.25%)

Thus, these narrow confidence intervals suggest trajectory prevalence will be estimated at a high level of precision.

Planned use of linked data and benefit to patients in England and Wales

STUDIES 2 AND 3

Linked deprivation status data will be used to assess the differences in durations of work absences due to musculoskeletal and/or mental health conditions from 2010-2021 (study 2) between levels of deprivation, and to assess association of deprivation with derived trajectories of work absences due to musculoskeletal and/or mental health conditions (study 3). Deprivation status could be a potentially important profile characteristic to help determine which trajectories of work absence particular patients are likely to follow, and thus this information may be very beneficial to GPs during initial consultation for a sickness absence, to help them provide more targeted and timely treatment to such patients.

Definition of the study population

ALL STUDIES

For all studies the source population will be:

• Patients aged 16 years and over
• Patients aged no greater than the current UK pension age (66 years)

And patients that have:

• A recorded fit note between 2010 and 2021
• A musculoskeletal or mental health coded consultation* within +/- 2 weeks of their first recorded fit note
• At least 2 years prior registration at their practice

• No previous recorded fit note (due to any reason)
• Registered with a practice in England
* identified using a pre-specified list of relevant codes developed for our previous studies

Index date will be the date of first recorded fit note.

STUDY 1
The population will be the subgroup receiving a first fit note in each calendar year from 2010-2021*.
*Note: the patterns in issuance of fit notes in 2020 and 2021 are likely to have been affected by the COVID-19 pandemic, for example it might be expected that there is higher fit note issuance due to mental health reasons, and fewer fit notes for other reasons (due to employers being more likely to accept self-certified absences of up to 14 days).

STUDY 2 and 3
The population will be a random sample of the subgroup receiving a first fit note in 2016-2018.

Selection of comparison group(s) or controls
This is not applicable to the three studies.

Exposures, outcomes and covariates
All of our studies will involve covariates, these will be measured 2 years before the index date, except for smoking and body mass index, which will be measured up to 5 years before index date (with the most recent value used), due to a greater amount of missing data expected for these covariates.

Musculoskeletal (MSK) conditions will be defined using the same Read/SNOMED codes used in the MSKCOM Keele based study utilising Aurum (ref 20_000105), this defines MSK as: osteoarthritis, inflammatory MSK, and/or the most common regional pain (back, knee, hip, and hand/wrist).

Mental health (MH) conditions will be defined using the same Read/SNOMED codes used in the MSKCOM Keele based study (ref 20_000105), this defines MH as: depression, anxiety, and/or stress.

In all three studies work absence will be defined as the first recorded fit note, with duration defined as the number of days between the issue date and the fit note end date (in instances where the fit note end date is not recorded, an approximation will be made using the median end date from patients that do have this data present). To differentiate between multiple fit note episodes for the same patient, a rule will be applied of there being at least a six-month gap between fit notes for it to be considered as a new episode.

STUDY 1
Covariates
• Age, sex, geographic region.
Outcomes

• Rates and duration of work absence due to musculoskeletal and/or mental health conditions (2010-2021)

STUDY 2
Covariates (note: these will only be used to assess differences in duration of work absence, separately for patients with first fit note due to: musculoskeletal conditions, mental health conditions, or both musculoskeletal and mental health conditions)
• Age, sex, geographic region, deprivation status (quintiles to be used), smoking, body mass index.
• Previous, current and follow-up consultation patterns for musculoskeletal and mental health conditions. Current consultations will compare by the type of MSK/MH that is linked to the start of the index fit note episode. Previous consultation patterns will be the number of MSK, MH, or MSK and MH (combined) consultations in the 2 years prior to index date, whilst follow-up consultation patterns will assess this over the 3-year follow-up period of each trajectory cohort.
• Comorbidity, defined as: Charlson comorbidity score, polypharmacy – number of different drugs, a count of consultations over the 2 years prior to index date, and the top two sickness certificate reasons with the highest sickness absence rates (excluding MSK and MH): injury/poisoning and respiratory conditions (25).
• Treatment received, defined as: analgesia (with hierarchies used ranging from basic analgesia to strong opioids, based on previous Keele studies), anti-inflammatory medication, and anti-depressants.

Outcomes
• Trajectories of work absence in patients with a recorded musculoskeletal and/or mental health consultation receiving a first fit note between 2016 and 2018 (with the focus on derivation of trajectories over the following three years).
• Duration of work absence due to musculoskeletal and/or mental health conditions (2016-2018)

STUDY 3
Covariates
• Age, sex, geographic region, deprivation status (quintiles to be used), smoking, body mass index.
• Previous, current and follow-up consultation patterns for musculoskeletal and mental health conditions. Current consultations will compare by the type of MSK/MH that is linked to the start of the index fit note episode. Previous consultation patterns will be the number of MSK, MH, or MSK and MH (combined) consultations in the 2 years prior to index date, whilst follow-up consultation patterns will assess this over the 3-year follow-up period of each trajectory cohort.
• Comorbidity, defined as: Charlson comorbidity score, polypharmacy – number of different drugs, a count of consultations over the 2 years prior to index date, and the top two sickness certificate reasons with the highest sickness absence rates (excluding MSK and MH): injury/poisoning and respiratory conditions (25).
• Treatment received, defined as: analgesia (with hierarchies used ranging from basic analgesia to strong opioids, based on previous Keele studies), anti-inflammatory medication, and anti-depressants.

Outcomes

• Trajectories of work absence in patients with a recorded musculoskeletal and/or mental health consultation receiving a first fit note between 2016 and 2018 (with the focus on association of covariates with derived trajectories from study 2).

Data/statistical analysis

STUDY 1

The annual incidence rate of fit note issuance due to musculoskeletal conditions (MSK), mental health (MH) conditions, or both musculoskeletal and mental health conditions will be determined, between 2010-2021, as will the percentage of total fit notes issued by type of fit note ('not fit for work', or one of the 'may be fit for work' options: 'a phased return to work', 'amended duties', 'altered hours', or 'workplace adaptations').

For each patient, the length of fit note will be determined based on:
1) Number of days for fit note within value field*
2) Where the above value is not present, time until end of fit note episode (defined using median of number of days of sickness absence from patients of same age and gender and reason for fit note (musculoskeletal, mental health, or both) that do have a recorded end date)
* This may not be present in all cases, and will depend on if the GP entered a value or not

The median (IQR) length of fit note by year will be reported.

Separately for MSK conditions, MH conditions, and both MSK and MH conditions, incidence rates will be stratified by: age, sex, geographic region.

When approximating the fit note end date for patients missing this information, in case the median fit note is being estimated from a low number of similar patients that do have this information recorded, a sensitivity analysis will be performed using 2 weeks as the fit note duration.

STUDY 2

Trajectories of work absence will be derived in patients with a recorded musculoskeletal and/or mental health consultation +/- 2 weeks of their first recorded fit note, that is received between 1st January 2016 and 31st December 2018 (with each year used as a separate cohort), then trajectories will be calculated using the subsequent three years' data as follow-up. As the number of included patients is expected to be large, and in order to avoid incurring significant computational problems, a random sample of 50,000 patients having a fit note due to musculoskeletal conditions, and 50,000 patients having a fit note due mental health conditions, will be selected for each yearly cohort. However, for the patients that have a first fit note due to both musculoskeletal and mental health conditions, as this patient group is expected to be small, all such patients will be used in trajectory analysis for each yearly cohort (2016-2018).

Simple methods of modelling trajectories, such as cumulative duration of work absence, will be contrasted against more complex approaches, such as different types of latent class analysis, including latent class growth analysis, growth mixture modelling, and latent transition analysis. Further trajectory methods may also be potentially discovered from the results of our systematic review. Different period intervals within the three

year follow-up will be tested for best operational performance, for example, splitting the data into 6-month recurring intervals, as in the work by Strauss et al (26).

For each trajectory method, clusters will be identified that represent groups of patients that are similar in their work absence patterns over follow-up. To keep the number of clusters to a manageable level, and maintain adequate power in the analysis, there will be a requirement that at least 5% of the study population is present in each cluster. The optimal number of clusters will be assessed using the Bayesian Information Criterion (BIC) and Akaike information criterion (AIC), with lower values being better. Average posterior class probability and entropy will also be considered, with values of > 0.7 indicating better fit. Finally, the trajectory clusters will be checked to ensure that they are also clinically meaningful.

Prevalence of each identified trajectory will be presented, and comparisons will be made between details of the trajectories (for example, number, and description of clusters) and corresponding prevalence across the different modelling methods, with similarities and differences critiqued.

A final model will be chosen, separately, for reason of index fit note due to: musculoskeletal conditions, mental health conditions, and both musculoskeletal and mental health conditions (if data is sufficient). Model choice will be based on best performance across BIC and AIC, average posterior class probability and entropy, and most clinically meaningful clusters derived.

The effect of using different measures of work absence on the trajectories will also be assessed, for example, absent yes/no, compared to cumulative number of days absent.

As in study 1, a sensitivity analysis will be performed considering 2 weeks to determine fit note end date for patients without this information (in instances where there is a low number of similar patients that do have this information recorded).

Furthermore, in this study, median (IQR) duration of (first) fit notes due to musculoskeletal and/or mental health conditions, for each yearly cohort (2016-2018) will be presented and stratified (separately for MSK conditions, MH conditions, and both MSK and MH conditions) by: sociodemographic characteristics (age, sex, geographic region, and deprivation status), health characteristics (body mass index, smoking, prior and current consultation patterns for musculoskeletal and mental health conditions, as well as consultation patterns during the follow-up period itself), comorbidity, and treatment received.

STUDY 3

Multivariable multinomial regression analyses will be performed, with each of the derived trajectories (under the chosen best performing model for each of musculoskeletal and mental health conditions, and possibly a model that combines musculoskeletal and mental health conditions) from study 2 as the outcome, and associations of trajectories with the following covariates will be assessed: sociodemographic characteristics (age, sex, geographic region, and deprivation status), health characteristics (body mass index, smoking, prior and current consultation patterns for musculoskeletal and mental health conditions, as well as consultation

patterns during the follow-up period itself), comorbidity, and treatment received. The model will account for clustering of patients in general practices and uncertainty in cluster membership.

Odds ratios of association for each covariate with trajectory membership will be presented, along with corresponding 95% confidence intervals to display the levels of uncertainty in these estimates.

Plan for addressing confounding
We have tried to be thorough as possible in pre-specifying as many important measured confounders as possible in all our multivariable analyses described in the data/statistical analysis section, but we acknowledge there will be some unobserved confounders (such as type of work) due to data access limitations.

Plans for addressing missing data
If no fit note, or consultation for a musculoskeletal or mental health condition is recorded, then we will assume by default that no fit note was issued, and that no such consultation occurred. For instances where there is a missing value for the end date of the fit note, we have made an assumption to use the median fit note duration from similar patients that do have a recorded end date. Furthermore, for any covariate that contains missing values (for example, possibly including: smoking, or body mass index), when testing for differences in incidence rates and duration of work absence, and association with trajectories, multiple imputation will be conducted in Stata (as long as the missing at random assumption is satisfied).

Patient or user group involvement
A patient and public involvement and engagement (PPIE) meeting will occur post-analysis. This PPIE meeting will allow us to gain feedback from patients and GPs to check that the different patient profiles that we discover through different work absence trajectories make sense to them, and this will help ensure that our findings are clinically relevant.

Plans for disseminating and communicating study results
The work resulting from our 3 studies will be presented at national conferences and a peer-reviewed publication is anticipated for each study. This work will inform future research exploring trajectories of work absence due to musculoskeletal and/or mental health conditions, especially in an English population.

Conflict of interest statement
There are no conflicts of interest to declare from study members.

Limitations of the study design, data sources, and analytic methods
The main study limitations are:

• The reason for a fit note being issued is not recorded in electronic health records, therefore we have had to make an assumption around this, that the reason for fit note issuance can be attributed to a musculoskeletal and/or mental health condition if the respective consultation for these conditions occurred within 2 weeks of the first ever fit

note. However, we acknowledge that this assumption does not allow for the issuance reason of subsequent fit notes to change over time.

• Not all fit notes have a recorded end date, thus in this instance we have assumed the end date to be based on the median fit note duration from patients that do have a recorded fit note end date (and with a sensitivity analysis based on fit note duration being 2 weeks, where number of patients with recorded fit note end dates is small).

• Unmeasured confounders - whilst we tried to be exhaustive in the factors we explored for differences in rates and duration of work absence (our studies 1 and 2), as well as associations of derived trajectories (our study 3), considerations around work-related factors would also have been very useful. For example, considerations around: the type of work (unskilled, manual, skilled etc), level of job control, emotional demands, interpersonal relations, and self-efficacy, are all important factors that are often noted to predict work absence. However, as such data is not available in medical records, we were limited by this lack of data access. There may also be other unobserved confounders that we have not adjusted our analyses for, due to lack of data access.

References

1. Henderson M, Clark C, Stansfeld S, Hotopf M. A lifecourse approach to long-term sickness absence-a Cohort study. PLoS One. 2012;7(5):e36645.
2. Health and Safety Executive. Health and safety at work: Summary statistics for Great Britain 2020 [Internet]. 2020.
3. Wynne-Jones G, Cowen J, Jordan JL, Uthman O, Main CJ, Glozier N. Absence from work and return to work in people with back pain: A systematic review and meta-analysis. Occup Environ Med. 2014;71(6):448–58.
4. Waddell G, Burton KA. Is work good for your health and well-being? 2006.
5. Department for Work and Pensions. Guidance for GPs, employers, hospital doctors and occupational health practitioners on using fit notes (statements of fitness for work) [Internet]. 2018 [cited 2021 Aug 5].
6. Jung T, Wickrama KAS. An Introduction to Latent Class Growth Analysis and Growth Mixture Modeling. Soc Personal Psychol Compass. 2008 Jan 1;2(1):302–17.
7. Schultz IZ, Crook J, Meloche GR, Berkowitz J, Milner R, Zuberbier OA, et al. Psychosocial factors predictive of occupational low back disability: Towards development of a return-to-work model. Pain. 2004 Jan;107(1–2):77–85.
8. Storheim K, Brox JI, Holm I, Bo K. Predictors of return to work in patients sick listed for sub-acute low back pain: a 12-month follow-up study. J Rehabil Med. 2005;37(6):365–71.
9. Reiso H, Nygard JF, Jorgensen GS, Holanger R, Soldal D, Bruusgaard D. Back to work: predictors of return to work among patients with back disorders certified as sick: a two-year follow-up study. Spine (Phila Pa 1976). 2003;28(13):1464–8.
10. Pedersen P, Lund T, Lindholdt L, Nohr EA, Jensen C, Søgaard HJ, et al. Labour market trajectories following sickness absence due to self-reported all cause morbidity--a longitudinal study. BMC Public Health. 2016 Apr 16;16(1):1–10.
11. Kerbo HR. Social Stratification and Inequality: Class Conflict in Historical, Comparative, and Global Perspective. McGraw-Hill; 2003.
12. Department of Health and Social Security. Inequalities in Health: Report of a Research Working Group. HM Stationery Office; 1980.

13. Gray AM. Inequalities in Health. The Black Report: A Summary and Comment. Int J Heal Serv. 1982 Jul 1;12(3):349–80.
14. Marmot MG, Stansfeld S, Patel C, North F, Head J, White I, et al. Health inequalities among British civil servants: the Whitehall II study. Lancet. 1991;337(8754):1387–93.
15. Marmot M, Allen J, Goldblatt P, Boyce T, McNeish D, Grady M, et al. Fair Society, Healthy Lives - The Marmot Review: Strategic Review of Health Inequalities in England post-2010. 2010.
16. Kasl S V, Jones BA. The impact of job loss and retirement on health. In: Social epidemiology. Oxford University Press New York; 2000. p. 118–36.
17. Bartley M. Health Inequality: An Introduction to Concepts, Theories and Methods. 2nd ed. John Wiley & Sons; 2016. 264 p.
18. Thomas C, Benzeval M, Stansfeld SA. Employment transitions and mental health: An analysis from the British household panel survey. J Epidemiol Community Health. 2005 Mar;59(3):243–9.
19. Waddell G, Burton KA, Kendall NA. Vocational Rehabilitation, what works, for whom and when? 2008.
20. Department for Work and Pensions. Getting the most out of the fit note: GP guidance. 2015.
21. Welsh VK, Mallen CD, Wynne-Jones G, Jinks C. Exploration of GPs' views and use of the fit note: a qualitative study in primary care. Br J Gen Pract. 2012 May 1;62(598):e363–70.
22. Parsonage M, Saini G. Mental health at work: The business costs ten years on [Internet]. 2017.
23. HSE. Work related musculoskeletal disorder statistics (WRMSDs) in Great Britain, 2020 [Internet]. London; 2020.
24. NHS Digital. Fit Notes Issued by GP Practices, England March 2021 [Internet]. 2021.
25. Wynne-Jones G, Mallen CD, Mottram S, Main CJ, Dunn KM. Identification of UK sickness certification rates, standardised for age and sex. Br J Gen Pract. 2009 Jul 1;59(564):510–6.
26. Strauss VY, Jones PW, Kadam UT, Jordan KP. Distinct trajectories of multimorbidity in primary care were identified using latent class growth analysis. J Clin Epidemiol. 2014/07/23. 2014 Oct;67(10):1163–71.

List of appendices

Table 1. The 27 Medical Code IDs identified that relate to fit notes (used for feasibility count process)

Medical Code ID	Term	Read Code
1653961000000113	MED3 (2010) issued by hand, may be fit for work	9D1A
1769641000006111	MED3 (2010) certificate issued - recommend phased return to work	9D1E
1769661000006110	MED3 (2010) certificate issued - recommend amended duties	9D1G

115910000000114	Med3 certification status	9D1Z
115610000000115	Med3 certificate issued to patient	9D11
176962100000006116	MED3 (2010) certificate issued to patient	9D1C
176965100000006113	MED3 (2010) certificate issued - recommend altered hours	9D1F
165335100000000114	eMED3 (2010) new statement issued, not fit for work	9D15
176967100000006115	MED3 (2010) certificate issued - recommend workplace adaptation	9D1H
165392100000000117	MED3 (2010) issued by hand, not fit for work	9D19
115510000000118	MED3 - doctor's statement	9D1
342010000000114	Med3 certificate issued - back to work	9D14
165366100000000118	eMED3 (2010) new statement issued, may be fit for work	9D16
114856100000000113	MED5 statement requested	9D23
115649100000000117	MED5 certificate requested	9D23-1
116410000000117	MED5 - issued to patient	9D21-z306
116210000000112	MED5 - doctor's special statement	9D2
116610000000116	MED5 status	9D2Z
116310000000114	MED5 issued to patient	9D21
735851000000113	Benefits agency reports unfit for work	13JJ0
250873012	Unfit for work	13JJ
250932010	Time off work	13JX
796806100000006117	Med3 certificate issued to patient	^ESCTME796806
124878810000006115	MED3 issued to patient	^ESCT1248788
124878710000006118	MED3 issued - back to work	^ESCT1248787
799947100000006114	Med3 certificate issued - back to work	^ESCTME799947
453682100000006110	Amount of time off work	^ESCTAM453682

Grant ID:{Empty}

Appendix F

Amended data request protocol, submitted to CPRD via eRAP online system on

12/09/2023, and approved on 20/09/2023

General information

Protocol reference Id
21_000665

Study title
Longitudinal trajectories of work absence in patients with musculoskeletal and, or, mental health conditions

Research area
Health Services Delivery, Methodological

Does this protocol describe an observational study using purely CPRD data?
Yes

Does this protocol involve requesting any additional information from GPs, or contact with patients?
No

Research team

Role: Chief Investigator
Title: Reader in Epidemiology and Clinical Trials
Full name: Gwenllian Wynne-Jones
Affiliation/organization: Keele University
Email: g.wynne-jones@keele.ac.uk
Will this person be analysing the data?: No
Status: Confirmed

Role: Corresponding Applicant
Title: PhD Student
Full name: Amardeep Legha
Affiliation/organization: Keele University
Email: a.s.legha@keele.ac.uk
Will this person be analysing the data?: Yes
Status: Confirmed

Role: Collaborator
Title: Data Manager
Full name: James Bailey
Affiliation/organization: Keele University

Email: j.bailey4@keele.ac.uk
Will this person be analysing the data?: Yes
Status: Confirmed

Role: Collaborator
Title: Professor
Full name: Clare Holdsworth
Affiliation/organization: Keele University
Email: c.m.holdsworth@keele.ac.uk
Will this person be analysing the data?: No
Status: Confirmed

Role: Collaborator
Title: Professor of Biostatistics
Full name: Kelvin Jordan
Affiliation/organization: Keele University
Email: k.p.jordan@keele.ac.uk
Will this person be analysing the data?: Yes
Status: Confirmed

Role: Collaborator
Title: NIHR Clinical Lecturer in General Practice
Full name: Victoria Welsh
Affiliation/organization: Keele University
Email: v.welsh@keele.ac.uk
Will this person be analysing the data?: No
Status: Confirmed

Access to data

Sponsor: Keele University

Funding source for the study
Is the funding source for the study the same as Chief Investigator's affiliation? No
Funding source for the study: Economic and Social Research Council (ESRC)

Institution conducting the research
Is the institution conducting the research the same as Chief Investigator's affiliation? Yes
Institution conducting the research: Keele University

Method to access the data
Indicate the method that will be used to access the data: Institutional multi-study licence
Is the institution the same as Chief Investigator's affiliation? Yes
Institution name: Keele University

Extraction by CPRD
Will the dataset be extracted by CPRD? No

Multiple data delivery
This study requires multiple data extractions over its lifespan: No

Data processors
Data processor is: Same as the chief investigator's affiliation
Processing: Yes
Accessing: Yes
Storing: Yes
Processing area: UK

Information on data

Primary care data: CPRD Aurum

Do you require data linkages? No

Patient level data: {Empty}

NCRAS data: {Empty}

Covid 19 linkages: {Empty}

Area level data
Do you require area level data?: Yes
Practice level (UK): {Empty}
Patient level (England only): Patient Level Index of Multiple Deprivation

Withheld concepts
Are withheld concepts required?: No

Linkage to a dataset not listed
Are you requesting a linkage to a dataset not listed? No

Patient data privacy
Does any person named in this application already have access to any of these data in a patient identifiable form, or associated with an identifiable patient index? No

Protocol information

Lay summary
Aches and pains, as well as mental health conditions, are one of the biggest causes of work absence. Most people return-to-work reasonably quickly after an episode of healthcare, but around 10 in 100 go on to have a longer-term absence of more than 12 months.

If a sickness absence lasts more than 7 days and sick pay is required, a General Practitioner (GP) can issue a fit note, that contains their recommendations about a potential return to work. Fit note information is recorded in primary care electronic

health records. Having access to CPRD allows us to access fit note data, and use it as a measure of work absence, to investigate work absence patterns over time. This is important, because during a consultation with a GP it is often difficult to tell who is at risk of longer-term absence. So we plan to use fit note data in patients with pain or a mental health condition to see if we can find common patterns of work absence such as having a long absence or returning to work quickly.

We also want to see if the health and sociodemographic characteristics of a person affect their chances of following a particular work absence pattern. For example, are people living in more disadvantaged neighbourhoods more likely to have a longer-term work absence?

Ultimately, the goal is to allow GPs to give more specific support to their patients at first consultation, to aid the return-to-work process.

Technical summary
BACKGROUND

Ability to work is one of the biggest drivers of social inequalities, leading to adverse health and social outcomes. Absence from work due to musculoskeletal and/or mental health conditions accounts for the majority of healthcare costs and productivity losses. Most people return-to-work relatively quickly following an episode of healthcare, but approximately 10% go on to have a longer-term work absence of > 12 months.

Fit notes are statements issued by GPs that record their medical recommendations regarding a potential return-to-work for patients absent for more than 7 days. They are recorded in primary care electronic health records; access to such data potentially allows uncovering of patterns of work absence over time (trajectories). Knowledge of these trajectories and associated characteristics could help GPs better distinguish patients at higher risk of sustained long-term work absence during initial consultation, and thus potentially offer earlier and more targeted intervention to such patients.

AIMS AND OBJECTIVES

1) To derive, and compare using different statistical methods, common longitudinal trajectories of work absence as measured by receipt of fit notes, for a population consulting their GP with a musculoskeletal and/or mental health condition
2) To derive health and sociodemographic characteristics associated with these trajectories

METHODS

For a population absent from work due to musculoskeletal and/or mental health conditions:

Study 1: Derivation of rates and duration of work absence (2010-2021), with differences examined by: sociodemographic characteristics (age, sex, and geographic region).

Study 2: Derivation of trajectories of work absence (2016-2018); contrasted using simple methods of modelling trajectories, against more complex approaches (such as different types of latent class analysis).

Study 3: Multivariable multinomial regression analyses to test association of each derived trajectory with the sociodemographic characteristics specified in study 1, as well as deprivation status, health characteristics, comorbidity, and treatment received.

Outcomes to be measured
STUDY 1
Rates and duration of work absence

STUDY 2
Trajectories of work absence (derivation)

STUDY 3
Trajectories of work absence (association of characteristics)

Objectives, specific aims and rationale
GENERAL OBJECTIVE
To describe the common longitudinal trajectories of work absence as measured by receipt of fit notes, in a population consulting their general practitioner with a musculoskeletal or mental health condition, and to derive profiles of patients within each of these trajectories. This information will be used to help inform a more timely and targeted intervention approach for GPs during first consultation with their patients, potentially making treatments more likely to succeed, and reducing the adverse health and social potential of worklessness.

SPECIFIC AIMS
Undertake the following in patients with a musculoskeletal and/or mental health condition in CPRD Aurum in England:

1) To derive, and compare using different statistical methods, common longitudinal trajectories of work absence. This will test the hypotheses around the existence, and possible identification in primary care, of different trajectories of work absence for patients with a musculoskeletal and/or mental health condition.
2) To derive health and sociodemographic characteristics associated with these trajectories. This will test the hypothesis that patients with certain health and sociodemographic characteristics, comorbidities, and receiving particular treatments, are more likely to belong to certain trajectories of work absence over others.

RATIONALE FOR STUDY
In the UK, it is known that long-term sickness absence is one of the main drivers of social inequalities, which lead to adverse social and economic outcomes for the individual, the employer, and wider society (1). In financial terms, for example, an estimated £16.2 billion was lost in the UK in the year 2018/19 due to work-related injury and ill health (2). Musculoskeletal and mental health conditions account for the majority of sickness absences in the UK, with for example, 27% and 55% of total working days lost due to ill health in 2019/20, due to musculoskeletal and mental health

conditions respectively (2). Although research shows that the majority of people do return to work within a short time frame following a medical problem, around 10% are expected to go on to experience longer term absences of > 12 months (3). The lengthier the work absence, the harder it is for the individual to return to work, and the more adverse health and social problems they experience (4). Work is generally good for physical and mental health, and worklessness is associated with poorer health and well-being (4).

If sickness absence persists for more than seven continuous days, a fit note can be issued. This is a written statement from a GP or other qualified medical practitioner that records the medical advice a patient has received regarding their fitness to work (5). Primary care electronic health records (EHRs), provide a valuable opportunity to examine work absence, by assessing patterns in the issuance of fit notes, through use of trajectory analysis. A trajectory describes the evolution of a repeated measure over time (for example, the course of work absence). This is achieved by detecting subgroups of individuals with similar patterns in a set of longitudinal heterogeneous data. Classification into subgroups is performed such that individuals share more similarities within their subgroup, than outside of the subgroup (6). During an initial GP consultation for sickness absence, it is challenging to determine which patients are at the highest risk of sustained long-term work absence; determining trajectories of work absence and the patient characteristics associated with them may assist with this. However, few studies of work absence trajectories have been conducted, especially for work absence due to musculoskeletal and/or mental health conditions (the two most common causes of work absence). Much of the work absence literature has been focused on investigating risk factors for experiencing a sickness absence; the few studies that do consider participants already on a sickness absence at baseline tend to consider only dichotomous outcomes, such as still having a sickness absence/no return-to-work after: 3 months (7), 12 months (8), or 2 years (9). Dichotomous outcomes may under- or over-estimate work absence (10); instead, using trajectories of work absence utilises the heterogeneity in both speed and duration of return-to-work spells, to form different trajectory clusters, and considers returning to work as a process as oppose to a fixed outcome.

To the best of our knowledge, our study will be the first to explore trajectories of work absence due to musculoskeletal and/or mental health conditions in an English population, as well as the profiles of patients associated with such trajectories. This is important because use of such trajectories can help GPs better understand, at initial consultation, which patients are more likely to undergo a detrimental course of work absence over time, and thus to act on this information by providing more timely and targeted support for such patients.

Study background

Long-term sickness absence in the UK is one of the main drivers of social inequalities, leading to adverse social and health outcomes (1). The costs of long-term sickness absence are significant, for example, an estimated £16.2 billion was lost in the UK in the year 2018/19 due to work-related injury and ill health (2).

Social inequality is the degree to which differences exist amongst groups in society, or "the condition where people have unequal access to valued resources, services, and

positions in the society" (11)(p.11). Much research has been conducted on social inequalities that influence health, especially since the 1980 publication of The Black Report (otherwise known as 'the Report of the Working Group on Inequalities in Health') (12). The authors demonstrated that since the inception of the National Health Service (NHS) in 1947, differences in risk of mortality had not reduced between different social classes, and these health inequalities had actually increased from the 1950s to the 1970s (13). The foremost finding from this report was that this was due to social inequalities rather than the NHS per se, due to differences that for example relate to: income, nutrition, education, and housing (13).

Other reports followed, such as the Whitehall II study which showed that social class (defined as employment grade) was inversely associated with risk of morbidity under a wide range of diseases, and with risk of mortality (14). The ground-breaking Strategic Review of Health Inequalities in England post-2010 (SRHIE, 2010) report also followed, and aimed to devise effective strategies to reduce health inequalities present in England (15). One of the foremost policy recommendations of the SRHIE (2010) concerns employment. Marmot et al set as a specific policy objective, the need to improve access to and quality of good employment, across all social strata. Additionally, they emphasized the need to reduce long-term unemployment across all social strata. This is important because it is known that work is generally good for physical and mental health, and worklessness is associated with poorer health and well-being (4).

Patterns of employment and work absence over time both reflect and reinforce the social gradient, and demonstrate the inequalities of access to labour market opportunities. People who are in lower socioeconomic positions are at a greater risk of unemployment (16), and being unemployed is associated with a greater rate of long-term illness (17), as well as mental illness (18). For the majority of the healthy working population the spiral towards worklessness tends to start with the onset of ill health, if this progresses to a point in which sick pay is required through the state, a fit note will usually need to be obtained from a medical practitioner, usually a GP. The premise behind a fit note is that the right kind of work is generally good for a patient's physical and mental well-being (4), and that it is not necessary to be completely fit to work in many instances (19); hence fit notes are administered with the ultimate goal of helping patients return to work as soon as they can and aiding their recovery (20)(21).

Although research shows that the majority of people do return to work within a short time frame following a medical problem, around 10% are expected to go on to experience longer term absences of > 12 months (3). The lengthier the work absence, the harder it is for the individual to return to work, and the more adverse health and social problems they experience (4). Poor health, and in particular chronic conditions, significantly affect one's ability to work. Chief among these are musculoskeletal and mental health conditions. For example, it was estimated that in the 2016/17 financial year £10.6 billion was spent on mental health related sickness absence in the UK (22). Additionally, it was shown that musculoskeletal disorders accounted for 27% of total working days lost due to ill health in 2019/20, and that 8.9 million working days were lost due to this condition in this time period (23).

Fit notes are recorded in primary care electronic health records. Having access to such data provides a unique opportunity to examine work absence, by assessing patterns in the issuance of fit notes over time (i.e., trajectories). During an initial GP consultation for sickness absence, it is challenging to determine which patients are at the highest risk of sustained long-term work absence; determining trajectories of work absence and the patient characteristics associated with them may assist with this.

Trajectories of work absence are scarcely studied. To the best of our knowledge, our study will be the first to explore trajectories of work absence due to musculoskeletal and/or mental health conditions in an English population, as well as the profiles of patients associated with such trajectories. This is important because a more complete understanding of the intricate courses of work absence over time for such individuals is hypothesised to be useful in tackling the issue of work absence. Hence, use of trajectories in this context can help GPs to better understand, at initial consultation, which patients are more likely to undergo a detrimental course of work absence over time, and thus to act on this information by providing more timely and targeted support for such patients.

Study type
Study aim 1 will be addressed through a descriptive study.

Study aim 2 will be addressed through a methodological/hypothesis generating study.

Study aim 3 will be addressed through a hypothesis generating study.

Study design
The research will involve three inter-related studies, each mapped to its own objective.

STUDY 1: will involve cross-sectional analyses to establish incidence rate and duration of work absence due to musculoskeletal and/or mental health conditions, and with trends in rates of absence over time to be compared, and differences explored by: sociodemographic characteristics (age, sex, and geographic region).

STUDY 2: will involve a retrospective cohort study whereby trajectories of work absence due to musculoskeletal and/or mental health conditions will be derived; there will also be a methodological component to this study as simple methods of modelling trajectories (such as cumulative duration of work absence), will be contrasted against more complex approaches (such as different types of latent class analysis, including latent class growth analysis, growth mixture modelling, and latent transition analysis).

STUDY 3: will involve a retrospective cohort study using multivariable multinomial regression analyses to test association of each derived trajectory in study 2, with: sociodemographic characteristics (age, sex, geographic region, and deprivation status), health characteristics (body mass index, smoking, and prior and current consultation patterns for musculoskeletal and mental health conditions), comorbidity (defined by Charlson Index, polypharmacy, and consultation count), and treatment received (defined as analgesia, anti-inflammatory medication, and anti-depressants).

Feasibility counts
In order to inform feasibility counts of the number of patients with a new fit note in a given time period, a list of Medical Code IDs was first devised using Read/SNOMED codes (provided in the Appendix).

Using this list of Read/SNOMED codes, and applying an estimate of the percentage of total fit notes that were issued due to musculoskeletal conditions (using data from a March 2021 NHS Digital Report) (24), feasibility counts (for patients with minimum age >= 16 years, and at least 2 years prior registration) suggest around 60,000 patients per year with a first fit note due to musculoskeletal conditions (2010-2014), up to around 120,000 per year (2019).

Similarly, feasibility counts suggest around 140,000 patients per year with a first fit note due to mental health conditions (2010-2014), up to around 300,000 per year (2019).

Sample size considerations
Based on the feasibility counts seen in the previous section, large numbers of patients are expected in our studies.

If, for example, there are 50,000 patients fulfilling our study 2 population criteria with a first fit note due to a musculoskeletal condition during 2016-2018, and 50,000 with a first fit note due to a mental health condition, then precision of trajectory prevalence can be estimated as the following (using a normal approximation to the binomial calculation with this sample size):
• 10% trajectory prevalence will be (9.74%, 10.26)
• 30% trajectory prevalence will be (29.60%, 30.40%)
• 50% trajectory prevalence will be (49.56%, 50.44%)

Thus, these narrow 95% confidence intervals suggest trajectory prevalence will be estimated at a high level of precision.

Planned use of linked data and benefit to patients in England and Wales
STUDIES 2 AND 3
Linked deprivation status data will be used to assess the differences in durations of work absences due to musculoskeletal and/or mental health conditions from 2010-2021 (study 2) between levels of deprivation, and to assess association of deprivation with derived trajectories of work absences due to musculoskeletal and/or mental health conditions (study 3). Deprivation status could be a potentially important profile characteristic to help determine which trajectories of work absence particular patients are likely to follow, and thus this information may be very beneficial to GPs during initial consultation for a sickness absence, to help them provide more targeted and timely treatment to such patients.

Definition of the study population
ALL STUDIES
For all studies the source population will be:
• Patients aged 16 years and over
• Patients aged no greater than the current UK pension age (66 years)
And patients that have:

• A recorded fit note between 2010 and 2021
• A musculoskeletal or mental health coded consultation* in the 2 weeks prior to their first recorded fit note
• At least 2 years prior registration at their practice
• No previous recorded fit note (due to any reason)
• Registered with a practice in England
* identified using a pre-specified list of relevant codes developed for our previous studies

Index date will be the date of first recorded fit note.

STUDY 1
The population will be the subgroup receiving a first fit note in each calendar year from 2010-2021*.
*Note: the patterns in issuance of fit notes in 2020 and 2021 are likely to have been affected by the COVID-19 pandemic, for example it might be expected that there is higher fit note issuance due to mental health reasons, and fewer fit notes for other reasons (due to employers being more likely to accept self-certified absences of up to 14 days).

STUDY 2 and 3
The population will be the subgroup receiving a first fit note in 2016-2018.

Selection of comparison groups/controls
This is not applicable to the three studies.

Exposures, outcomes and covariates
All of our studies will involve covariates, these will be measured 2 years before the index date, except for smoking and body mass index, which will be measured up to 5 years before index date (with the most recent value used), due to a greater amount of missing data expected for these covariates.

Musculoskeletal (MSK) conditions will be defined using the same Read/SNOMED codes used in the MSKCOM Keele based study utilising Aurum (ref 20_000105), this defines MSK as: osteoarthritis, inflammatory MSK, and/or the most common regional pain (back, knee, hip, and hand/wrist).

Mental health (MH) conditions will be defined using the same Read/SNOMED codes used in the MSKCOM Keele based study (ref 20_000105), this defines MH as: depression, anxiety, and/or stress.

In all three studies work absence will be defined as the first recorded fit note, with duration defined as the number of days between the issue date and the fit note end date (in instances where the fit note end date is not recorded, an approximation will be made using the median end date from patients that do have this data present). To differentiate between multiple fit note episodes for the same patient, a rule will be applied of there being at least a six-month gap between fit notes for it to be considered as a new episode.

STUDY 1

Covariates
• Age, sex, geographic region.
Outcomes
• Rates and duration of work absence due to musculoskeletal and/or mental health conditions (2010-2021)

STUDY 2
Covariates (note: these will only be used to assess differences in duration of work absence, separately for patients with first fit note due to: musculoskeletal conditions, mental health conditions, or both musculoskeletal and mental health conditions)
• Age, sex, geographic region, deprivation status (quintiles to be used), smoking, body mass index.
• Previous, current and follow-up consultation patterns for musculoskeletal and mental health conditions. Current consultations will compare by the type of MSK/MH that is linked to the start of the index fit note episode. Previous consultation patterns will be the number of MSK, MH, or MSK and MH (combined) consultations in the 2 years prior to index date, whilst follow-up consultation patterns will assess this over the 3-year follow-up period of each trajectory cohort.
• Comorbidity, defined as: Charlson comorbidity score, polypharmacy – number of different drugs, and a count of consultations over the 2 years prior to index date.
• Treatment received, defined as: analgesia (with hierarchies used ranging from basic analgesia to strong opioids, based on previous Keele studies), anti-inflammatory medication, and anti-depressants.

Outcomes
• Trajectories of work absence in patients with a recorded musculoskeletal and/or mental health consultation receiving a first fit note between 2016 and 2018 (with the focus on derivation of trajectories over the following three years).
• Duration of work absence due to musculoskeletal and/or mental health conditions (2016-2018)

STUDY 3
Covariates
• Age, sex, geographic region, deprivation status (quintiles to be used), smoking, body mass index.
• Previous, current and follow-up consultation patterns for musculoskeletal and mental health conditions. Current consultations will compare by the type of MSK/MH that is linked to the start of the index fit note episode. Previous consultation patterns will be the number of MSK, MH, or MSK and MH (combined) consultations in the 2 years prior to index date, whilst follow-up consultation patterns will assess this over the 3-year follow-up period of each trajectory cohort.
• Comorbidity, defined as: Charlson comorbidity score, polypharmacy – number of different drugs, and a count of consultations over the 2 years prior to index date.
• Treatment received, defined as: analgesia (with hierarchies used ranging from basic analgesia to strong opioids, based on previous Keele studies), anti-inflammatory medication, and anti-depressants.

Outcomes

• Trajectories of work absence in patients with a recorded musculoskeletal and/or mental health consultation receiving a first fit note between 2016 and 2018 (with the focus on association of covariates with derived trajectories from study 2).

Data/statistical analysis

STUDY 1

The annual incidence rate of fit note issuance due to musculoskeletal conditions (MSK), mental health (MH) conditions, or both musculoskeletal and mental health conditions will be determined, between 2010-2021, as will the percentage of total fit notes issued by type of fit note ('not fit for work', or one of the 'may be fit for work' options: 'a phased return to work', 'amended duties', 'altered hours', or 'workplace adaptations').

For each patient, the length of fit note will be determined based on:
1) Number of days for fit note within value field*
2) Where the above value is not present, time until end of fit note episode (defined using median of number of days of sickness absence from patients of same age and gender and reason for fit note (musculoskeletal, mental health, or both) that do have a recorded end date)
* This may not be present in all cases, and will depend on if the GP entered a value or not

The median (IQR) length of fit note by year will be reported.

Separately for MSK conditions, MH conditions, and both MSK and MH conditions, incidence rates will be stratified by: age, sex, geographic region.

When approximating the fit note end date for patients missing this information, in case the median fit note is being estimated from a low number of similar patients that do have this information recorded, a sensitivity analysis will be performed using 2 weeks as the fit note duration.

STUDY 2

Trajectories of work absence will be derived in patients with a recorded musculoskeletal and/or mental health consultation in the 2 weeks prior to their first recorded fit note, that is received between 1st January 2016 and 31st December 2018, then trajectories will be calculated using the subsequent three years' data as follow-up.

Simple methods of modelling trajectories, such as cumulative duration of work absence, will be contrasted against more complex approaches, such as different types of latent class analysis, including latent class growth analysis, growth mixture modelling, and latent transition analysis. Further trajectory methods may also be potentially discovered from the results of our systematic review. Different period intervals within the three year follow-up will be tested for best operational performance, for example, splitting the data into 6-month recurring intervals, as in the work by Strauss et al (26).

For each trajectory method, clusters will be identified that represent groups of patients that are similar in their work absence patterns over follow-up. To keep the number of clusters to a manageable level, and maintain adequate power in the analysis, there will be a requirement that at least 1% of the study population is present in each cluster. The

optimal number of clusters will be assessed using the Bayesian Information Criterion (BIC) and Akaike information criterion (AIC), with lower values being better. Average posterior class probability and entropy will also be considered, with values of > 0.7 indicating better fit. Finally, the trajectory clusters will be checked to ensure that they are also clinically meaningful.

Prevalence of each identified trajectory will be presented, and comparisons will be made between details of the trajectories (for example, number, and description of clusters) and corresponding prevalence across the different modelling methods, with similarities and differences critiqued.

A final model will be chosen, separately, for reason of index fit note due to: musculoskeletal conditions, mental health conditions, and both musculoskeletal and mental health conditions (if data is sufficient). Model choice will be based on best performance across BIC and AIC, average posterior class probability and entropy, and most clinically meaningful clusters derived.

The effect of using different measures of work absence on the trajectories will also be assessed, for example, absent yes/no, compared to cumulative number of days absent.

As in study 1, a sensitivity analysis will be performed considering 2 weeks to determine fit note end date for patients without this information (in instances where there is a low number of similar patients that do have this information recorded).

Furthermore, in this study, median (IQR) duration of (first) fit notes due to musculoskeletal and/or mental health conditions, for each yearly cohort (2016-2018) will be presented and stratified (separately for MSK conditions, MH conditions, and both MSK and MH conditions) by: sociodemographic characteristics (age, sex, geographic region, and deprivation status), health characteristics (body mass index, smoking, prior and current consultation patterns for musculoskeletal and mental health conditions, as well as consultation patterns during the follow-up period itself), comorbidity, and treatment received.

STUDY 3

Multivariable multinomial regression analyses will be performed, with each of the derived trajectories (under the chosen best performing model for each of musculoskeletal and mental health conditions, and possibly a model that combines musculoskeletal and mental health conditions) from study 2 as the outcome, and associations of trajectories with the following covariates will be assessed: sociodemographic characteristics (age, sex, geographic region, and deprivation status), health characteristics (body mass index, smoking, prior and current consultation patterns for musculoskeletal and mental health conditions, as well as consultation patterns during the follow-up period itself), comorbidity, and treatment received. The model will account for clustering of patients in general practices and uncertainty in cluster membership.

Odds ratios of association for each covariate with trajectory membership will be presented, along with corresponding 95% confidence intervals to display the levels of uncertainty in these estimates.

Plan for addressing confounding
We have tried to be thorough as possible in pre-specifying as many important measured confounders as possible in all our multivariable analyses described in the data/statistical analysis section, but we acknowledge there will be some unobserved confounders (such as type of work) due to data access limitations.

Plans for addressing missing data
If no fit note, or consultation for a musculoskeletal or mental health condition is recorded, then we will assume by default that no fit note was issued, and that no such consultation occurred. For instances where there is a missing value for the end date of the fit note, we have made an assumption to use the median fit note duration from similar patients that do have a recorded end date. Furthermore, for any covariate that contains missing values (for example, possibly including: smoking, or body mass index), when testing for differences in incidence rates and duration of work absence, and association with trajectories, multiple imputation will be conducted in Stata (as long as the missing at random assumption is satisfied).

Patient or user group involvement
A patient and public involvement and engagement (PPIE) meeting will occur post-analysis. This PPIE meeting will allow us to gain feedback from patients and GPs to check that the different patient profiles that we discover through different work absence trajectories make sense to them, and this will help ensure that our findings are clinically relevant.

Plans for disseminating & communicating
The work resulting from our 3 studies will be presented at national conferences and a peer-reviewed publication is anticipated for each study. This work will inform future research exploring trajectories of work absence due to musculoskeletal and/or mental health conditions, especially in an English population.

Conflict of interest statement
There are no conflicts of interest to declare from study members.

Limitations of the study design
The main study limitations are:

• The reason for a fit note being issued is not recorded in electronic health records, therefore we have had to make an assumption around this, that the reason for fit note issuance can be attributed to a musculoskeletal and/or mental health condition if the respective consultation for these conditions occurred within the 2 weeks prior to the first ever fit note. However, we acknowledge that this assumption does not allow for the issuance reason of subsequent fit notes to change over time.

• Not all fit notes have a recorded end date, thus in this instance we have assumed the end date to be based on the median fit note duration from patients that do have a recorded fit note end date (and with a sensitivity analysis based on fit note duration being 2 weeks, where number of patients with recorded fit note end dates is small).

• Unmeasured confounders - whilst we tried to be exhaustive in the factors we explored for differences in rates and duration of work absence (our studies 1 and 2), as well as associations of derived trajectories (our study 3), considerations around work-related factors would also have been very useful. For example, considerations around: the type of work (unskilled, manual, skilled etc), level of job control, emotional demands, interpersonal relations, and self-efficacy, are all important factors that are often noted to predict work absence. However, as such data is not available in medical records, we were limited by this lack of data access. There may also be other unobserved confounders that we have not adjusted our analyses for, due to lack of data access.

References

1. Henderson M, Clark C, Stansfeld S, Hotopf M. A lifecourse approach to long-term sickness absence-a Cohort study. PLoS One. 2012;7(5):e36645.
2. Health and Safety Executive. Health and safety at work: Summary statistics for Great Britain 2020 [Internet]. 2020.
3. Wynne-Jones G, Cowen J, Jordan JL, Uthman O, Main CJ, Glozier N. Absence from work and return to work in people with back pain: A systematic review and meta-analysis. Occup Environ Med. 2014;71(6):448–58.
4. Waddell G, Burton KA. Is work good for your health and well-being? 2006.
5. Department for Work and Pensions. Guidance for GPs, employers, hospital doctors and occupational health practitioners on using fit notes (statements of fitness for work) [Internet]. 2018 [cited 2021 Aug 5].
6. Jung T, Wickrama KAS. An Introduction to Latent Class Growth Analysis and Growth Mixture Modeling. Soc Personal Psychol Compass. 2008 Jan 1;2(1):302–17.
7. Schultz IZ, Crook J, Meloche GR, Berkowitz J, Milner R, Zuberbier OA, et al. Psychosocial factors predictive of occupational low back disability: Towards development of a return-to-work model. Pain. 2004 Jan;107(1–2):77–85.
8. Storheim K, Brox JI, Holm I, Bo K. Predictors of return to work in patients sick listed for sub-acute low back pain: a 12-month follow-up study. J Rehabil Med. 2005;37(6):365–71.
9. Reiso H, Nygard JF, Jorgensen GS, Holanger R, Soldal D, Bruusgaard D. Back to work: predictors of return to work among patients with back disorders certified as sick: a two-year follow-up study. Spine (Phila Pa 1976). 2003;28(13):1464–8.
10. Pedersen P, Lund T, Lindholdt L, Nohr EA, Jensen C, Søgaard HJ, et al. Labour market trajectories following sickness absence due to self-reported all cause morbidity--a longitudinal study. BMC Public Health. 2016 Apr 16;16(1):1–10.
11. Kerbo HR. Social Stratification and Inequality: Class Conflict in Historical, Comparative, and Global Perspective. McGraw-Hill; 2003.
12. Department of Health and Social Security. Inequalities in Health: Report of a Research Working Group. HM Stationery Office; 1980.
13. Gray AM. Inequalities in Health. The Black Report: A Summary and Comment. Int J Heal Serv. 1982 Jul 1;12(3):349–80.
14. Marmot MG, Stansfeld S, Patel C, North F, Head J, White I, et al. Health inequalities among British civil servants: the Whitehall II study. Lancet. 1991;337(8754):1387–93.
15. Marmot M, Allen J, Goldblatt P, Boyce T, McNeish D, Grady M, et al. Fair Society, Healthy Lives - The Marmot Review: Strategic Review of Health Inequalities in England post-2010. 2010.

16. Kasl S V, Jones BA. The impact of job loss and retirement on health. In: Social epidemiology. Oxford University Press New York; 2000. p. 118–36.
17. Bartley M. Health Inequality: An Introduction to Concepts, Theories and Methods. 2nd ed. John Wiley & Sons; 2016. 264 p.
18. Thomas C, Benzeval M, Stansfeld SA. Employment transitions and mental health: An analysis from the British household panel survey. J Epidemiol Community Health. 2005 Mar;59(3):243–9.
19. Waddell G, Burton KA, Kendall NA. Vocational Rehabilitation, what works, for whom and when? 2008.
20. Department for Work and Pensions. Getting the most out of the fit note: GP guidance. 2015.
21. Welsh VK, Mallen CD, Wynne-Jones G, Jinks C. Exploration of GPs' views and use of the fit note: a qualitative study in primary care. Br J Gen Pract. 2012 May 1;62(598):e363–70.
22. Parsonage M, Saini G. Mental health at work: The business costs ten years on [Internet]. 2017.
23. HSE. Work related musculoskeletal disorder statistics (WRMSDs) in Great Britain, 2020 [Internet]. London; 2020.
24. NHS Digital. Fit Notes Issued by GP Practices, England March 2021 [Internet]. 2021.
25. Strauss VY, Jones PW, Kadam UT, Jordan KP. Distinct trajectories of multimorbidity in primary care were identified using latent class growth analysis. J Clin Epidemiol. 2014/07/23. 2014 Oct;67(10):1163–71.

List of appendices

Table 1. The 27 Medical Code IDs identified that relate to fit notes (used for feasibility count process)

Medical Code ID	Term	Read Code
1653961000000113	MED3 (2010) issued by hand, may be fit for work	9D1A
1769641000006111	MED3 (2010) certificate issued - recommend phased return to work	9D1E
1769661000006110	MED3 (2010) certificate issued - recommend amended duties	9D1G
1159100000114	Med3 certification status	9D1Z
1156100000115	Med3 certificate issued to patient	9D11
1769621000006116	MED3 (2010) certificate issued to patient	9D1C
1769651000006113	MED3 (2010) certificate issued - recommend altered hours	9D1F
1653351000000114	eMED3 (2010) new statement issued, not fit for work	9D15

176967100000 06115	MED3 (2010) certificate issued - recommend workplace adaptation	9D1H
165392100000 00117	MED3 (2010) issued by hand, not fit for work	9D19
115510000000 118	MED3 - doctor's statement	9D1
342010000000 114	Med3 certificate issued - back to work	9D14
165366100000 00118	eMED3 (2010) new statement issued, may be fit for work	9D16
114856100000 00113	MED5 statement requested	9D23
115649100000 00117	MED5 certificate requested	9D23-1
116410000000 117	MED5 - issued to patient	9D21-z306
116210000000 112	MED5 - doctor's special statement	9D2
116610000000 116	MED5 status	9D2Z
116310000000 114	MED5 issued to patient	9D21
735851000000 0113	Benefits agency reports unfit for work	13JJ0
250873012	Unfit for work	13JJ
250932010	Time off work	13JX
796806100000 06117	Med3 certificate issued to patient	^ESCTME 796806
124878810000 06115	MED3 issued to patient	^ESCT124 8788
124878710000 06118	MED3 issued - back to work	^ESCT124 8787
799947100000 06114	Med3 certificate issued - back to work	^ESCTME 799947
453682100000 06110	Amount of time off work	^ESCTAM 453682

Grant ID:{Empty}

Appendix G

The Guidelines for Reporting on Latent Trajectory Studies (GRoLTS) Checklist (2016)[173]

Table G.1. Items of the GRoLTS Checklist (2016)[173]

Item Number	Checklist Item
1	Is the metric of time used in the statistical model reported?
2	Is information presented about the mean and variance of time within a wave?
3a.	Is the missing data mechanism reported?
3b.	Is a description provided of what variables are related to attrition/missing data?
3c.	Is a description provided of how missing data in the analyses were dealt with?
4	Is information about the distribution of the observed variables included?
5	Is the software mentioned?
6a.	Are alternative specifications of within-class heterogeneity considered (e.g., LCGA vs. GMM) and clearly documented? If not, was sufficient justification provided as to eliminate certain specifications from consideration?
6b.	Are alternative specifications of the between-class differences in variance-covariance matrix structure considered and clearly documented? If not, was sufficient justification provided as to eliminate certain specifications from consideration?
7	Are alternative shape/functional forms of the trajectories described?
8	If covariates have been used, can analyses still be replicated?
9	Is information reported about the number of random start values and final iterations included?
10	Are the model comparison (and selection) tools described from a statistical perspective?

11	Are the total number of fitted models reported, including a one-class solution?
12	Are the number of cases per class reported for each model (absolute sample size, or proportion)?
13	If classification of cases in a trajectory is the goal, is entropy reported?
14a.	Is a plot included with the estimated mean trajectories of the final solution?
14b.	Are plots included with the estimated mean trajectories for each model?
14c.	Is a plot included of the combination of estimated means of the final model and the observed individual trajectories split out for each latent class?
15	Are characteristics of the final class solution numerically described (i.e., means, SD/SE, n, CI, etc.)?
16	Are the syntax files available (either in the appendix, supplementary materials, or from the authors)?

Abbreviations: LCGA = Latent Class Growth Analysis; GMM = Growth Mixture Modelling; SD = Standard Deviation; SE = Standard Error; CI = Confidence Interval; n = number of individuals

Appendix H

CPRD Read Codes for Study Outcomes and Covariates

Table H.1. Old Set of 39 Medical Code IDs used to identify fit notes in CPRD (for feasibility count process)

Medical Code ID	Term	Read Code
1653961000000113	MED3 (2010) issued by hand, may be fit for work	9D1A
1769611000006112	MED3 (2010) certificate not issued to patient	9D1B
1769631000006118	MED3 (2010) certificate duplicate issued	9D1D
1769641000006111	MED3 (2010) certificate issued - recommend phased return to work	9D1E
1769661000006110	MED3 (2010) certificate issued - recommend amended duties	9D1G
11591000000114	Med3 certification status	9D1Z
11561000000115	Med3 certificate issued to patient	9D11
1653701000000112	eMED3 (2010) duplicate issued, not fit for work	9D17
11581000000112	Med3 certificate not issued to patient	9D13
1769621000006116	MED3 (2010) certificate issued to patient	9D1C
1769651000006113	MED3 (2010) certificate issued - recommend altered hours	9D1F
1653351000000114	eMED3 (2010) new statement issued, not fit for work	9D15
11571000000110	Med3 certificate duplicate issued	9D12
1769671000006115	MED3 (2010) certificate issued - recommend workplace adaptation	9D1H
1653741000000110	eMED3 (2010) duplicate issued, may be fit for work	9D18
1653921000000117	MED3 (2010) issued by hand, not fit for work	9D19
11551000000118	MED3 - doctor's statement	9D1

34201000000114	Med3 certificate issued - back to work	9D14
1653661000000118	eMED3 (2010) new statement issued, may be fit for work	9D16
1148561000000113	MED5 statement requested	9D23
1156491000000117	MED5 certificate requested	9D23-1
11641000000117	MED5 - issued to patient	9D21-z306
11621000000112	MED5 - doctor's special statement	9D2
11661000000116	MED5 status	9D2Z
11631000000114	MED5 issued to patient	9D21
11651000000119	MED5 - not able to issue	9D22
735851000000113	Benefits agency reports unfit for work	13JJ0
1921301000006115	Benefits agency reports unfit for work but fit note no longer needed	13JJ1
250873012	Unfit for work	13JJ
250932010	Time off work	13JX
7968061000006117	Med3 certificate issued to patient	^ESCTME796806
12487881000006115	MED3 issued to patient	^ESCT1248788
12487891000006117	MED3 not issued to patient	^ESCT1248789
7968151000006115	Med3 certificate not issued to patient	^ESCTME796815
12487861000006113	MED3 duplicate issued	^ESCT1248786
7968111000006116	Med3 certificate duplicate issued	^ESCTME796811
12487871000006118	MED3 issued - back to work	^ESCT1248787
7999471000006114	Med3 certificate issued - back to work	^ESCTME799947
4536821000006110	Amount of time off work	^ESCTAM453682

Table H.2. Final Set of 27 Medical Code IDs used to identify fit notes in CPRD

Medical Code ID	Term	Read Code
1653961000000113	MED3 (2010) issued by hand, may be fit for work	9D1A
1769641000006111	MED3 (2010) certificate issued - recommend phased return to work	9D1E
1769661000006110	MED3 (2010) certificate issued - recommend amended duties	9D1G
11591000000114	Med3 certification status	9D1Z
11561000000115	Med3 certificate issued to patient	9D11
1769621000006116	MED3 (2010) certificate issued to patient	9D1C
1769651000006113	MED3 (2010) certificate issued - recommend altered hours	9D1F
1653351000000114	eMED3 (2010) new statement issued, not fit for work	9D15
1769671000006115	MED3 (2010) certificate issued - recommend workplace adaptation	9D1H
1653921000000117	MED3 (2010) issued by hand, not fit for work	9D19
11551000000118	MED3 - doctor's statement	9D1
34201000000114	Med3 certificate issued - back to work	9D14
1653661000000118	eMED3 (2010) new statement issued, may be fit for work	9D16
1148561000000113	MED5 statement requested	9D23
1156491000000117	MED5 certificate requested	9D23-1
11641000000117	MED5 - issued to patient	9D21-z306
11621000000112	MED5 - doctor's special statement	9D2
11661000000116	MED5 status	9D2Z
11631000000114	MED5 issued to patient	9D21

735851000000113	Benefits agency reports unfit for work	13JJ0
250873012	Unfit for work	13JJ
250932010	Time off work	13JX
7968061000006117	Med3 certificate issued to patient	^ESCTME796806
12487881000006115	MED3 issued to patient	^ESCT1248788
12487871000006118	MED3 issued - back to work	^ESCT1248787
7999471000006114	Med3 certificate issued - back to work	^ESCTME799947
4536821000006110	Amount of time off work	^ESCTAM453682

Appendix I

Investigating Spread of Issued Fit Notes for 2016 MSK Cohort, by Absence Episode Number and Fit Note Number

Figure I.1. Distribution of 2016 MSK Cohort Fit Notes by Absence Episode Number

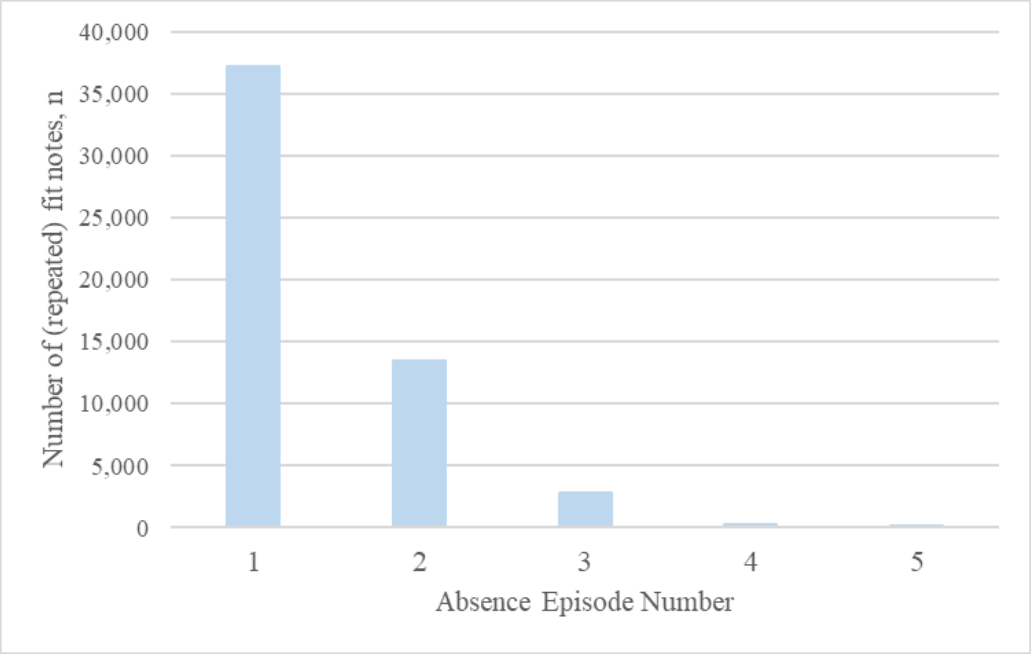
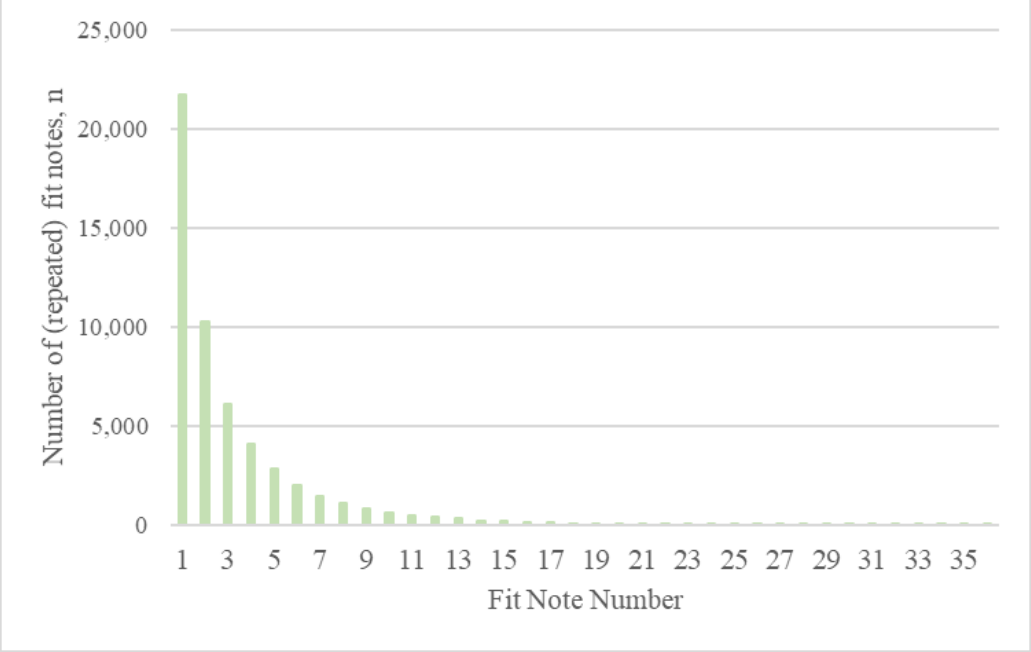


Figure I.2. Distribution of 2016 MSK Cohort Fit Notes by Fit Note Number



Appendix J

Model Fit Indices and Plots for LGCMs in Trajectory Derivation Analyses

Table J.1. Model Fit Indices of LGCMs for MSK 2016-2018 Cohort

	Interval Approach					
	1[a]	**2[b]**	**3[c]**	**4[d] (Non-Piecewise)**	**4[d] (Piecewise)**	**5[e]**
n	42,905	43,130	42,222	42,905	42,905	43,130
RMSEA Estimate (90% CI)	0.059 (0.053, 0.065)	0.039 (0.036, 0.041)	0.046 (0.044, 0.049)	0.048 (0.047, 0.050)	0.041 (0.039, 0.042)	0.046 (0.044, 0.047)
CFI	0.99	0.99	0.94	0.94	0.96	0.94
TLI	0.96	0.98	0.92	0.93	0.95	0.94
SRMR	0.03	0.04	0.05	0.07	0.05	0.07

Abbreviations: n = number of individuals; LGCM = Latent Growth Curve Model; MSK = Musculoskeletal; RMSEA = Root Mean Square Error of Approximation; CI = Confidence Interval; CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; SRMR = Standardised Root Mean Residual

[a] Three Monthly Intervals, Year One Data Only

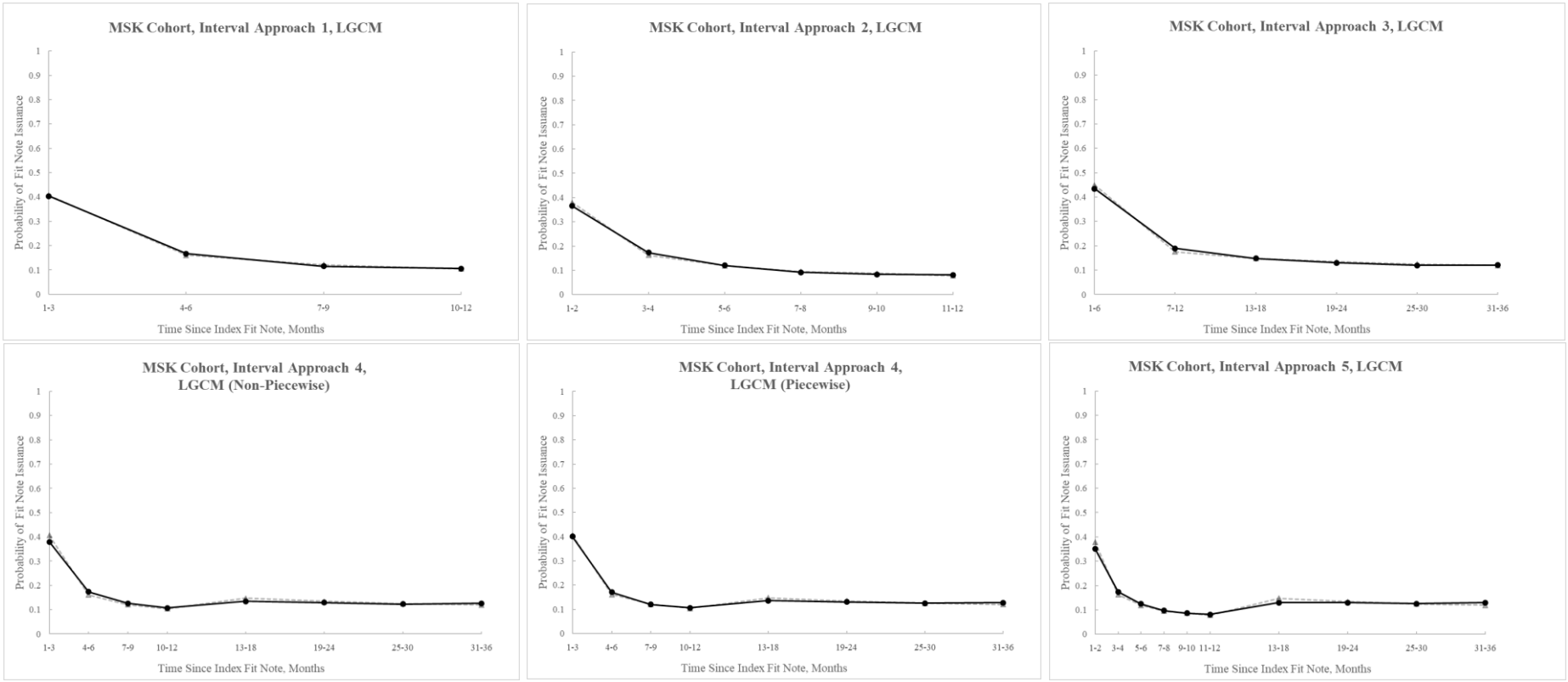
[b] Two Monthly Intervals, Year One Data Only

[c] Six Monthly Intervals, Years One to Three Data

[d] Three Monthly Intervals (Year One); Six Monthly (Years Two to Three)

[e] Two Monthly Intervals (Year One); Six Monthly (Years Two to Three)

Figure J.1. LGCM Trajectory Plots for MSK 2016-2018 Cohort



Abbreviations: LGCM = Latent Growth Curve Model; MSK = Musculoskeletal.
The solid lines represent the probability of fit note issuance in the given time interval (model estimated data), whilst the dotted lines represent the proportion of individuals issued a fit note (observed data).

Table J.2. Model Fit Indices of LGCMs for MH 2016-2018 Cohort

	Interval Approach					
	1[a]	**2[b]**	**3[c]**	**4[d] (Non-Piecewise)**	**4[d] (Piecewise)**	**5[e]**
n	61,900	62,355	60,536	61,900	61,900	62,355
RMSEA Estimate (90% CI)	0.072 (0.068, 0.077)	0.053 (0.051, 0.055)	0.061 (0.059, 0.063)	0.066 (0.065, 0.067)	0.051 (0.049, 0.052)	0.059 (0.058, 0.060)
CFI	0.98	0.98	0.91	0.90	0.95	0.93
TLI	0.95	0.97	0.88	0.89	0.93	0.92
SRMR	0.03	0.05	0.07	0.09	0.05	0.09

Abbreviations: LGCM = Latent Growth Curve Model; MH = Mental Health; RMSEA = Root Mean Square Error of Approximation; CI = Confidence Interval; CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; SRMR = Standardised Root Mean Residual

[a] Three Monthly Intervals, Year One Data Only

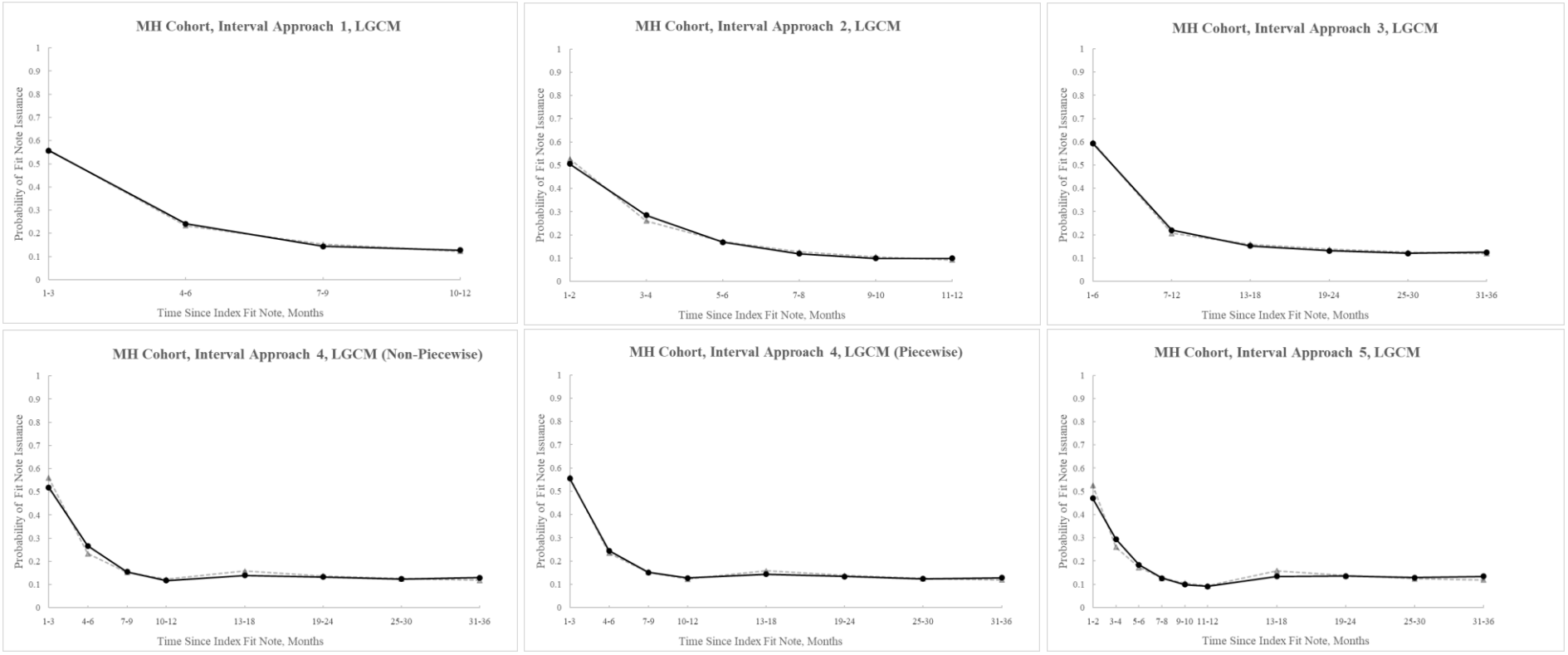
[b] Two Monthly Intervals, Year One Data Only

[c] Six Monthly Intervals, Years One to Three Data

[d] Three Monthly Intervals (Year One); Six Monthly (Years Two to Three)

[e] Two Monthly Intervals (Year One); Six Monthly (Years Two to Three)

Figure J.2. LGCM Trajectory Plots for MH 2016-2018 Cohort



Abbreviations: LGCM = Latent Growth Curve Model; MH = Mental Health.
The solid lines represent the probability of fit note issuance in the given time interval (model estimated data), whilst the dotted lines represent the proportion of individuals issued a fit note (observed data).

Appendix K

Assessing the Individual Variability of the 'Intermittent Low' Trajectory Class for the Optimal Five-Class LCGA Models Under Interval Approach 2

Table K.1. Summary of Observed Fit Note Issuance Patterns During Year One Follow-Up for the 'Intermittent Low' Trajectory Class in the Optimal Five-Class LCGA Model for the Incident MSK Condition Fit Note Cohort

Outcome Pattern (2 monthly)[a]	n (%)[b]	Median Intervals With Fit Note Received
0,1,0,0,0,0	905 (12.44%)	2
0,0,1,0,0,0	677 (9.31%)	
0,0,0,1,0,0	578 (7.95%)	
0,0,0,0,1,0	565 (7.77%)	
1,0,1,0,0,0	391 (5.38%)	
1,0,0,0,0,1	374 (5.14%)	
1,0,0,0,1,0	342 (4.70%)	
1,0,0,1,0,0	271 (3.73%)	
0,1,1,0,0,0	261 (3.59%)	
0,0,0,0,1,1	206 (2.83%)	
1,0,1,1,0,0	151 (2.08%)	
0,0,1,1,0,0	147 (2.02%)	
1,0,0,0,1,1	146 (2.01%)	
0,0,0,1,1,0	141 (1.94%)	
1,1,0,1,0,0	130 (1.79%)	
0,1,1,1,0,0	122 (1.68%)	
1,0,0,1,1,0	99 (1.36%)	
1,1,0,0,0,1	92 (1.27%)	
1,1,0,0,1,0	88 (1.21%)	
0,0,0,1,1,1	81 (1.11%)	
0,1,0,1,0,0	79 (1.09%)	
0,1,0,0,1,0	68 (0.94%)	
1,0,1,1,1,0	68 (0.94%)	
1,1,0,1,1,0	68 (0.94%)	
1,1,1,0,0,1	61 (0.84%)	
1,1,0,0,1,1	60 (0.83%)	
1,0,0,1,1,1	59 (0.81%)	
0,1,0,0,0,1	58 (0.8%)	

0,0,1,0,1,0	57 (0.78%)
0,0,1,0,0,1	56 (0.77%)
0,0,1,1,1,1	54 (0.74%)
0,1,1,0,1,0	54 (0.74%)
0,1,1,1,1,0	51 (0.70%)
0,0,0,1,0,1	47 (0.65%)
0,0,1,1,1,0	46 (0.63%)
1,0,0,1,0,1	41 (0.56%)
1,0,1,0,1,0	40 (0.55%)
0,1,0,1,1,0	35 (0.48%)
0,1,1,0,1,1	35 (0.48%)
1,0,1,1,0,1	35 (0.48%)
0,1,1,1,0,1	33 (0.45%)
1,1,0,1,0,1	30 (0.41%)
0,0,1,0,1,1	27 (0.37%)
0,1,.,.,.,.	27 (0.37%)
1,0,1,0,1,1	27 (0.37%)
0,1,0,1,1,1	25 (0.34%)
1,0,1,0,0,1	25 (0.34%)
0,1,0,0,1,1	19 (0.26%)
0,1,0,.,.,.	18 (0.25%)
0,1,1,0,0,1	17 (0.23%)
0,0,1,0,.,.	12 (0.17%)
0,0,1,1,0,1	12 (0.17%)
0,1,0,0,.,.	12 (0.17%)
0,1,0,0,0,.	12 (0.17%)
0,0,0,0,1,.	11 (0.15%)
0,1,0,1,0,1	11 (0.15%)
0,0,0,1,.,.	10 (0.14%)
0,1,1,.,.,.	9 (0.12%)
0,0,1,.,.,.	8 (0.11%)
1,0,1,.,.,.	8 (0.11%)
1,0,1,0,.,.	8 (0.11%)
0,0,0,1,0,.	7 (0.10%)
0,0,1,0,0,.	7 (0.10%)
0,1,1,1,.,.	6 (0.08%)
1,0,0,0,1,.	6 (0.08%)
1,0,0,1,.,.	6 (0.08%)

Abbreviations: LCGA = Latent Class Growth Analysis; MSK = Musculoskeletal.

[a] A value of 1 is used to denote that at least one fit note was issued in the given time interval, 0 denotes that no fit notes were issued, and a period indicates that the individual had missing data in the time interval.

[b] Cell counts less than five are not shown in accordance with CPRD reporting guidelines (to reduce risk of patient identification).
The pattern is chronologically ordered in two monthly time intervals during the first year of follow-up since index fit note. For example, a pattern of 1,0,0,0,0,0 indicates that a fit note was issued in the first two months since index fit note, but not in the ensuing 10 months.
Note: trajectory prevalence is based on most likely latent class membership, not posterior probabilities, as individuals are treated as whole persons in the observed data, hence the posterior probabilities cannot be used (as explained earlier in Section 5.2.6).

Table K.2. Summary of Observed Fit Note Issuance Patterns During Year One Follow-Up for the 'Intermittent Low' Trajectory Class in the Optimal Five-Class LCGA Model for the Incident MH Condition Fit Note Cohort

Outcome Pattern (2 monthly)[a]	n (%)[b]	Median Intervals With Fit Note Received
0,1,0,0,0,0	1295 (11.95%)	
1,0,1,0,0,0	682 (6.29%)	
0,0,1,0,0,0	662 (6.11%)	
1,0,0,0,0,1	655 (6.05%)	
0,0,0,1,0,0	621 (5.73%)	
0,0,0,0,1,0	598 (5.52%)	
0,1,1,0,0,0	476 (4.39%)	
1,0,0,1,0,0	439 (4.05%)	
1,0,0,0,1,0	415 (3.83%)	
1,1,0,1,0,0	302 (2.79%)	
1,0,1,1,0,0	260 (2.4%)	
0,0,0,0,1,1	259 (2.39%)	
1,1,0,0,0,1	246 (2.27%)	
1,0,0,0,1,1	231 (2.13%)	
0,1,1,1,0,0	204 (1.88%)	
0,0,0,1,1,0	196 (1.81%)	
0,0,1,1,0,0	185 (1.71%)	
1,0,0,1,1,0	184 (1.7%)	2
1,1,0,0,1,0	165 (1.52%)	
1,1,0,1,1,0	146 (1.35%)	
1,1,1,0,0,1	141 (1.3%)	
1,0,1,1,1,0	136 (1.26%)	
1,0,0,1,1,1	123 (1.14%)	
1,1,0,0,1,1	118 (1.09%)	
0,1,0,1,0,0	115 (1.06%)	
0,1,1,1,1,0	110 (1.02%)	
0,0,0,1,1,1	106 (0.98%)	
0,0,1,1,1,0	89 (0.82%)	
0,1,0,0,0,1	87 (0.8%)	
0,0,0,1,0,1	81 (0.75%)	
0,1,1,0,1,0	72 (0.66%)	
0,0,1,0,1,0	69 (0.64%)	
0,0,1,1,1,1	69 (0.64%)	
1,0,1,0,1,0	69 (0.64%)	
0,1,0,1,1,0	67 (0.62%)	

0,1,1,0,1,1	66 (0.61%)
0,1,0,0,1,0	64 (0.59%)
1,0,1,1,0,1	62 (0.57%)
0,0,1,0,0,1	58 (0.54%)
0,1,0,1,1,1	56 (0.52%)
1,0,1,0,0,1	55 (0.51%)
1,1,0,1,0,1	53 (0.49%)
1,0,1,0,1,1	51 (0.47%)
1,0,0,1,0,1	50 (0.46%)
0,1,0,0,1,1	47 (0.43%)
0,1,.,.,.,.	45 (0.42%)
0,1,1,1,0,1	42 (0.39%)
0,1,1,0,0,1	39 (0.36%)
0,0,1,0,1,1	35 (0.32%)
0,1,0,1,0,1	33 (0.3%)
0,1,0,.,.,.	30 (0.28%)
0,0,1,1,0,1	28 (0.26%)
0,1,0,0,.,.	28 (0.26%)
0,0,1,.,.,.	25 (0.23%)
1,0,1,.,.,.	25 (0.23%)
0,0,0,1,.,.	20 (0.18%)
0,1,0,0,0,.	20 (0.18%)
0,1,1,.,.,.	20 (0.18%)
1,0,1,0,.,.	19 (0.18%)
1,0,1,0,0,.	13 (0.12%)
0,0,0,0,1,.	12 (0.11%)
0,1,1,1,.,.	12 (0.11%)
1,0,0,0,1,.	12 (0.11%)
1,0,0,1,.,.	12 (0.11%)
1,0,1,1,.,.	12 (0.11%)
0,0,1,0,0,.	11 (0.1%)
0,1,1,0,.,.	11 (0.1%)
1,0,0,1,0,.	11 (0.1%)
0,0,1,0,.,.	10 (0.09%)
0,0,0,1,0,.	9 (0.08%)
0,1,1,0,0,.	9 (0.08%)
0,0,1,1,.,.	8 (0.07%)
1,1,0,1,.,.	7 (0.06%)
1,0,1,1,0,.	6 (0.06%)

Abbreviations: LCGA = Latent Class Growth Analysis; MH = Mental Health.

[a] A value of 1 is used to denote that at least one fit note was issued in the given time interval, 0 denotes that no fit notes were issued, and a period indicates that the individual had missing data in the time interval.

[b] Cell counts less than five are not shown in accordance with CPRD reporting guidelines (to reduce risk of patient identification).
The pattern is chronologically ordered in two monthly time intervals during the first year of follow-up since index fit note. For example, a pattern of 1,0,0,0,0,0 indicates that a fit note was issued in the first two months since index fit note, but not in the ensuing 10 months.
Note: trajectory prevalence is based on most likely latent class membership, not posterior probabilities, as individuals are treated as whole persons in the observed data, hence the posterior probabilities cannot be used (as explained earlier in Section 5.2.6).

Appendix L

CPRD Read Codes for Identifying Type of MH Condition

Table L.1. The 189 Medical Code IDs used to identify type of MH condition in CPRD Aurum (for Study 3)

Medical Code ID	Term	Read Code	Type of MH Condition
388271000006110	[X]Grief reaction	Eu432-2	Depression
398351000006110	[X]Mild anxiety depression	Eu412-1	Anxiety and Depression
398561000006117	[X]Mixed anxiety and depressive disorder	Eu412	Anxiety and Depression
398851000006117	[X]Mood - affective disorders	Eu3	Anxiety and Depression
296245018	[X]Other mixed anxiety disorders	Eu413	Anxiety
296207019	[X]Other persistent mood affective disorders	Eu34y	Anxiety and Depression
419841000006116	[X]Persistant anxiety depression	Eu341-4	Anxiety and Depression
296208012	[X]Persistent mood affective disorder, unspecified	Eu34z	Anxiety and Depression
296204014	[X]Persistent mood affective disorders	Eu34	Anxiety and Depression
295478016	Acute situational disturbance	E2830	Anxiety and Depression
481850015	Agitated	1B16	Anxiety
474161000006117	Agitated - symptom	1B16-1	Anxiety
5024071000006110	Anxiety depression	^ESCTAN502407	Anxiety and Depression
488211000006112	Anxiety with depression	E2003	Anxiety and Depression
958591000006110	Grief	EMISCGR1	Depression
974331000006113	Grief NOS	EMISCGR4	Depression

123751014	Grief reaction	E2900	Depression
959811000006113	Irritable/agitated	EMISCIR1	Stress
981111000006112	Irritable/agitated/aggressive	EMISCIR3	Stress
1976491000006110	Mixed anxiety and depressive reaction	EMISICD10|F4322	Anxiety and Depression
1976371000006110	Other single mood affective disorders, mixed affective episode	EMISICD10|F3800	Anxiety and Depression
296249012	[X]Anxiety disorder, unspecified	Eu41z	Anxiety
363641000006114	[X]Anxiety hysteria	Eu41y-1	Anxiety
363651000006111	[X]Anxiety neurosis	Eu411-1	Anxiety
363661000006113	[X]Anxiety NOS	Eu41z-1	Anxiety
363671000006118	[X]Anxiety reaction	Eu411-2	Anxiety
363681000006115	[X]Anxiety state	Eu411-3	Anxiety
388071000006116	[X]Generalized anxiety disorder	Eu411	Anxiety
296238018	[X]Other anxiety disorders	Eu41	Anxiety
401881014	[X]Other specified anxiety disorders	Eu41y	Anxiety
418031000006113	[X]Panic attack	Eu410-1	Anxiety
296239014	[X]Panic disorder [episodic paroxysmal anxiety]	Eu410	Anxiety
418051000006118	[X]Panic disorder with agoraphobia	Eu400-2	Anxiety
418061000006116	[X]Panic state	Eu410-2	Anxiety
1230451012	Adjustment reaction with anxious mood	E2924	Anxiety
294992011	Agoraphobia with panic attacks	E2021	Anxiety
294963012	Anxiety state NOS	E200z	Anxiety
294953016	Anxiety state unspecified	E2000	Anxiety
488201000006114	Anxiety states	E200	Anxiety
516601000000113	Anxious	1B13-2	Anxiety
2536376012	Anxiousness	1B13	Anxiety
488251000006113	Anxiousness - symptom	1B13-1	Anxiety

1488717011	C/O - panic attack	1B1V	Anxiety
294960010	Chronic anxiety	E2004	Anxiety
6550101000006110	Complaining of panic attack	^ESCTCO655010	Anxiety
481154010	Generalised anxiety disorder	E2002	Anxiety
253620013	O/E - anxious	2258	Anxiety
2549895012	O/E - panic attack	225J	Anxiety
6910391000006110	On examination - panic attack	^ESCTON691039	Anxiety
339044013	Panic attack	E2001-1	Anxiety
1210253015	Panic disorder	E2001	Anxiety
1808521000006110	Panic disorder without agoraphobia	EMISNQPA155	Anxiety
853241000006119	Phobic anxiety	EGTON122	Anxiety
223641000000116	Phobic anxiety	E202-2	Anxiety
294961014	Recurrent anxiety	E2005	Anxiety
1808511000006110	Recurrent panic attacks	JHCRE17	Anxiety
441512015	Separation anxiety disorder	E2920	Anxiety
909681000006110	[RFC] Depression	HNGNQRF13	Depression
359121000006116	[X] Reactive depression NOS	Eu32z-4	Depression
376691000006116	[X]Depression NOS	Eu32z-1	Depression
376701000006116	[X]Depressive conduct disorder	Eu920	Depression
376711000006118	[X]Depressive disorder NOS	Eu32z-2	Depression
376721000006114	[X]Depressive episode	Eu32	Depression
401872015	[X]Depressive episode, unspecified	Eu32z	Depression
376741000006119	[X]Depressive neurosis	Eu341-1	Depression
379431000006113	[X]Dysthymia	Eu341	Depression
1715771000006110	[X]Major depression, mild	Eu325	Depression
1715181000006110	[X]Major depression, moderately severe	Eu326	Depression

396081000006116	[X]Major depression, recurrent without psychotic symptoms	Eu332-2	Depression
1715781000006110	[X]Major depression, severe without psychotic symptoms	Eu327	Depression
213641000000111	[X]Mild depression	Eu324	Depression
296137015	[X]Mild depressive episode	Eu320	Depression
11918531000006100	[X]Mild depressive episode	^ESCT1191853	Depression
296138013	[X]Moderate depressive episode	Eu321	Depression
11918561000006100	[X]Moderate depressive episode	^ESCT1191856	Depression
398841000006119	[X]Monopolar depression NOS	Eu33z-1	Depression
399961000006118	[X]Neurotic depression	Eu341-3	Depression
401871010	[X]Other depressive episodes	Eu32y	Depression
296199015	[X]Other recurrent depressive disorders	Eu33y	Depression
423611000006111	[X]Prolonged single episode of reactive depression	Eu32z-3	Depression
424531000006118	[X]Recurr depress disorder cur epi severe without psyc sympt	Eu332	Depression
424551000006113	[X]Recurr severe episodes/psychogenic depressive psychosis	Eu333-4	Depression
401873013	[X]Recurrent depressive disorder	Eu33	Depression
296180012	[X]Recurrent depressive disorder, current episode mild	Eu330	Depression
296181011	[X]Recurrent depressive disorder, current episode moderate	Eu331	Depression
401876017	[X]Recurrent depressive disorder, unspecified	Eu33z	Depression
424631000006119	[X]Recurrent episodes of depressive reaction	Eu33-1	Depression
424641000006112	[X]Recurrent episodes of psychogenic depression	Eu33-2	Depression
424651000006114	[X]Recurrent episodes of reactive depression	Eu33-3	Depression
425411000006110	[X]SAD - Seasonal affective disorder	Eu33-5	Depression
425751000006115	[X]Seasonal depressive disorder	Eu33-4	Depression
401866015	[X]Severe depressive episode without psychotic symptoms	Eu322	Depression
11921301000006100	[X]Severe depressive episode without psychotic symptoms	^ESCT1192130	Depression
223741000000112	[X]Single episode major depression w'out psychotic symptoms	Eu322-2	Depression

426911000006111	[X]Single episode of depressive reaction	Eu32-1	Depression
426941000006110	[X]Single episode of psychogenic depression	Eu32-2	Depression
426971000006119	[X]Single episode of reactive depression	Eu32-3	Depression
426981000006116	[X]Single episode of reactive depressive psychosis	Eu323-4	Depression
1806431000006110	Adjustment disorder with depressed mood	EMISNQAD51	Depression
525921000006119	Brief depressive reaction	E290	Depression
295494011	Brief depressive reaction NOS	E290z	Depression
407062014	C/O - feeling depressed	1B17-1	Depression
407066012	C/O - feeling unhappy	1B17-2	Depression
295537016	Chronic depression	E2B1	Depression
5532721000006110	Complaining of feeling depressed	^ESCTCO553272	Depression
2164006016	Depressed	1B17	Depression
3071801000006110	Depressed	^ESCTDE307180	Depression
2164005017	Depressed mood	1BT	Depression
882671000006112	Depression	E2B-98	Depression
1823881000006110	Depression confirmed	EMISNQDE36	Depression
882681000006110	Depression NOS	E2B-99	Depression
295535012	Depressive disorder NEC	E2B	Depression
12727931000006100	Depressive disorder NEC	^ESCT1272793	Depression
3071791000006110	Depressive illness	^ESCTDE307179	Depression
1494612017	Depressive symptoms	1B1U-1	Depression
3094941000006110	Major depression, single episode	^ESCTMA309494	Depression
3094951000006110	Major depressive disorder, single episode	^ESCTMA309495	Depression
6000691000006110	Mild depression	^ESCTMI600069	Depression
882811000006119	Mild depression	Eu320-99	Depression
1975991000006110	Mild depressive episode, with somatic syndrome	EMISICD10|F3201	Depression

1975981000006110	Mild depressive episode, without somatic syndrome	EMISICD10|F3200	Depression
6000711000006110	Moderate depression	^ESCTMO600071	Depression
882821000006110	Moderate depression	Eu321-99	Depression
1976051000006110	Moderate depressive episode, with somatic syndrome	EMISICD10|F3211	Depression
1976021000006110	Moderate depressive episode, without somatic syndrome	EMISICD10|F3210	Depression
675861000006113	Neurotic depression reactive type	E204	Depression
253619019	O/E - depressed	2257	Depression
1976411000006110	Other recurrent mood affective disorders, recurrent brief depressive disorder	EMISICD10|F3810	Depression
295536013	Postviral depression	E2B0	Depression
3153071000006110	Recurrent brief depressive disorder	^ESCTRE315307	Depression
294844012	Recurrent depression	E1137	Depression
1976231000006110	Recurrent depressive disorder, current episode mild, with somatic syndrome	EMISICD10|F3301	Depression
1976211000006110	Recurrent depressive disorder, current episode mild, without somatic syndrome	EMISICD10|F3300	Depression
1976271000006110	Recurrent depressive disorder, current episode moderate, with somatic syndrome	EMISICD10|F3311	Depression
1976251000006110	Recurrent depressive disorder, current episode moderate, without somatic syndrome	EMISICD10|F3310	Depression
182721000006111	Recurrent major depressive episode	E113	Depression
294845013	Recurrent major depressive episode NOS	E113z	Depression
294837015	Recurrent major depressive episodes, mild	E1131	Depression
294838013	Recurrent major depressive episodes, moderate	E1132	Depression
182771000006112	Recurrent major depressive episodes, severe, no psychosis	E1133	Depression
294836012	Recurrent major depressive episodes, unspecified	E1130	Depression
182801000006114	Recurrent major depressive episodes,partial/unspec remission	E1135	Depression
369982012	Seasonal affective disorder	E118	Depression
882831000006113	Severe depression	Eu322-99	Depression
6000721000006110	Severe depression	^ESCTSE600072	Depression

3087971000006110	Severe recurrent major depression without psychotic features	^ESCTSE308797	Depression
401766011	Single major depressive episode	E112	Depression
294832014	Single major depressive episode NOS	E112z	Depression
294825017	Single major depressive episode, mild	E1121	Depression
12451461000006100	Single major depressive episode, moderate	^ESCT1245146	Depression
294826016	Single major depressive episode, moderate	E1122	Depression
142541000006115	Single major depressive episode, severe, without psychosis	E1123	Depression
294824018	Single major depressive episode, unspecified	E1120	Depression
1488626018	Symptoms of depression	1B1U	Depression
972861000006118	** Traumatic stress - effects of overwhelming experience	TRIQQZZ42	Stress
302741000006114	[D]State of emotional shock and stress, unspecified	R00zW	Stress
909701000006113	[RFC] Post traumatic stress disorder	HNGNQRF15	Stress
1755921000006110	[X]Acute post-traumatic stress disorder follow military comb	Eu433	Stress
362401000006115	[X]Acute reaction to stress	Eu430-2	Stress
362441000006118	[X]Acute stress reaction	Eu430	Stress
1755931000006110	[X]Chron post-traumatic stress disorder follow military comb	Eu434	Stress
1755941000006110	[X]Delayed post-traumat stress disorder follow military comb	Eu435	Stress
296271019	[X]Other reactions to severe stress	Eu43y	Stress
423021000006114	[X]Post - traumatic stress disorder	Eu431	Stress
424471000006119	[X]Reaction to severe stress, and adjustment disorders	Eu43	Stress
296272014	[X]Reaction to severe stress, unspecified	Eu43z	Stress
318033015	[X]State of emotional shock and stress, unspecified	Ryu58	Stress
295475018	Acute fugue state due to acute stress reaction	E281	Stress
295474019	Acute panic state due to acute stress reaction	E280	Stress
459711000006119	Acute posttrauma stress state	E2831	Stress
500650019	Acute reaction to stress	E28	Stress

3592701000006110	Acute reaction to stress	^ESCTAC359270	Stress
3592651000006110	Acute stress disorder	^ESCTAC359265	Stress
401810015	Acute stress reaction NOS	E28z	Stress
295476017	Acute stupor state due to acute stress reaction	E282	Stress
882651000006119	Chron post-traum stress dis	E29y1-99	Stress
959221000006118	Difficulty managing stress	EMISCDI6	Stress
338135018	Feeling stressed	1B1T	Stress
295483012	Other acute stress reaction NOS	E283z	Stress
295477014	Other acute stress reactions	E283	Stress
295512014	Other post-traumatic stress disorder	E29y1	Stress
12730101000006100	Other post-traumatic stress disorder	^ESCT1273010	Stress
3268061000006110	Posttraumatic stress disorder	^ESCTPO326806	Stress
3268091000006110	Post-traumatic stress syndrome	^ESCTPO326809	Stress
3268081000006110	PTSD - Post-traumatic stress disorder	^ESCTPT326808	Stress
4766771000006110	Stress reaction causing mixed disturbance	^ESCTST476677	Stress
121871000006118	Stress reaction causing mixed disturbance of emotion/conduct	E284	Stress
121881000006115	Stress related problem	1B1L	Stress
4934751000006110	Stressed out	^ESCTST493475	Stress

Abbreviations: MH = Mental Health.

Appendix M

Description of Derivation of Body Mass Index Health Characteristic for Trajectory-Covariate Association Analysis

Body mass index (BMI) was explored as a health characteristic in Chapter 8 when assessing associations of characteristics with optimal absence trajectories. Recorded BMI data was identified from a Read/SNOMED code list (n=233) from the MSKCOM study[125] (code list publicly available from https://doi.org/10.21252/878s-x990). This was the most complex characteristic derived in this thesis and required a multi-faceted approach (in line with other Keele-based CPRD studies).

Firstly, the BMI code list used was categorised into: 'underweight', 'normal', 'overweight', 'obese', 'weight management' (individuals in this category were referred onto a weight management program due to being overweight or obese), as well as codes that corresponded to a continuous value of: height, weight, or BMI. Hence the BMI code list used related to a combination of categorical and continuous BMI data.

These codes were then searched for in the CPRD Aurum database, in the five years prior to (and including) index fit note date. A five-year search period was chosen, similar to the search for smoking data, as a relatively high level of missing BMI data was anticipated (compared to the other characteristics derived from CPRD in this thesis).

Next, data cleaning was performed to only retain appropriate continuous BMI data:

- Any continuous height values that were present were retained, provided that the corresponding Read codes matched those previously identified for height (or that the specific unit of measurements recorded in the data matched those

corresponding to the most common height measurements), and that the values were between 1 to 2.2 metres (or 100 to 220 centimetres; any heights retained within this range were then converted to a metre metric).

- Any continuous weight values that were present were retained, provided that the corresponding Read codes matched those previously identified for weight (or that the specific unit of measurements recorded in the data matched those corresponding to the most common weight measurements), and that the values were between 30 to 300 kilograms.
- Any continuous BMI values that were present were retained, provided that the corresponding Read codes matched those previously identified for continuous BMI (or that the specific unit of measurements recorded in the data matched those corresponding to the most common BMI measurements), and that the values were between 10 to 80.

Thus, to exclude any extreme outliers, assumptions were made in the above process around possible plausible ranges for adult heights, weights, and BMI values.

Then, for any individuals that did have appropriate continuous BMI value data present (according to the above criteria) within this five-year period, the most recent BMI value was retained. For individuals with missing continuous BMI data, but where a continuous height and weight value were both present, a continuous BMI value was computed. This was calculated as weight (in kilograms) divided by height (in metres) squared. This calculation was based on taking the most recent height and weight values for the participant (within the five-year period).

Then, these calculated continuous BMI values (either directly through a continuous BMI value, or indirectly through height and weight data) were categorised according to NICE guidelines:[202]

- 'Underweight' if $10 \leq BMI < 18.5$
- 'Normal' if $18.5 \leq BMI < 25$
- 'Overweight' if $25 \leq BMI < 30$
- 'Obese' if $30 \leq BMI < 80$

If the BMI values were outside of a plausible range (defined as $BMI < 10$ or $BMI \geq 80$ in this study), they were categorised as 'missing'.

Thus, the (direct or indirect) continuous BMI value data had now been extracted and transformed into the same format as the original categorial BMI data that was derived directly from the code list. Hence, the overall BMI category options were: 'underweight', 'normal', 'overweight', 'obese', 'weight management', or 'missing'.

Therefore, with all the available BMI data now cleansed and in a consistent format, the next step was to retain only the most recent BMI category data per participant. Then, de-duplication was handled by first exploring the data. No individuals who had duplicate data (with more than one BMI category recorded on the most recent date), had 'missing' as a BMI option. For individuals who had 'weight management' as one of their duplicate entries, this was the entry that was deleted first. Then, de-duplication was finalised by application of a hierarchy that assumed the lowest BMI option for a participant. For example, if a participant had a BMI data entry recorded as 'overweight', and another as 'normal', both on the same date, only the 'normal' BMI entry was retained.

Any individuals who did not have any BMI data present retrieved from this five-year search were also coded into the ‘missing’ BMI category, and this category then renamed ‘not recorded’. Furthermore, as the ‘weight management’ category was sparse (for example, only 1% of the index MH condition fit note cohort was contained in this category), this was also collapsed into the ‘not recorded’ category.

Finally, the ‘underweight’ category was also sparsely populated, and was therefore collapsed into the ‘normal’ category. Thus, the final BMI categories used in this study were: ‘underweight/normal’, ‘overweight’, ‘obese’, or ‘not recorded’.

Appendix N

Unadjusted Models to Assess Association of Characteristics with Optimal Trajectories of Work Absence

Table N.1. Characteristics Associated with Optimal Trajectories of Work Absence Due to a MSK Condition Using the 'Single' Trajectory Class as the Reference (Unadjusted Model)

	Trajectory Class			
	Chronic Sustained[a] n=1,261	**Chronic Fast Decreasing[a]** n=1,333	**Intermittent Low[a]** n=7,272	**Short Term[a]** n=11,154
Sex				
Male	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)
Female	1.09 (0.96, 1.24)	**1.32 (1.15, 1.51)**	**1.36 (1.27, 1.46)**	**1.11 (1.04, 1.17)**
Age				
16-25 years	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)
26-35 years	**1.48 (1.11, 1.99)**	1.09 (0.83, 1.43)	1.05 (0.93, 1.18)	**1.28 (1.17, 1.40)**
36-45 years	**1.92 (1.46, 2.54)**	1.28 (0.99, 1.65)	0.93 (0.83, 1.05)	**1.21 (1.11, 1.33)**
46-55 years	**2.67 (2.04, 3.49)**	**1.71 (1.33, 2.19)**	1.02 (0.91, 1.14)	**1.45 (1.32, 1.59)**
56-66 years	**4.73 (3.64, 6.13)**	**3.15 (2.47, 4.00)**	**1.42 (1.25, 1.62)**	**1.63 (1.47, 1.81)**
Region[b]				
South of England	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)
North of England	**1.61 (1.38, 1.88)**	**1.93 (1.65, 2.25)**	**1.13 (1.02, 1.24)**	**1.35 (1.26, 1.45)**
Middle of England	**1.51 (1.29, 1.78)**	**1.30 (1.08, 1.57)**	**1.15 (1.05, 1.27)**	**1.19 (1.11, 1.28)**

IMD[c]				
1	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)
2	1.08 (0.83, 1.41)	1.22 (0.93, 1.59)	1.16 (1.00, 1.35)	1.03 (0.93, 1.14)
3	1.24 (0.95, 1.61)	1.18 (0.91, 1.54)	**1.34 (1.17, 1.54)**	0.99 (0.90, 1.10)
4	**1.60 (1.25, 2.06)**	1.22 (0.94, 1.59)	**1.64 (1.44, 1.87)**	0.96 (0.87, 1.06)
5	**2.60 (2.06, 3.27)**	**1.79 (1.41, 2.27)**	**1.88 (1.65, 2.14)**	0.97 (0.89, 1.07)
Missing	1.20 (0.78, 1.87)	**0.39 (0.16, 0.99)**	0.82 (0.59, 1.16)	1.29 (1.10, 1.51)
MSK Consultations - Prior 2 Years[d]				
0	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)
1	**1.61 (1.37, 1.88)**	**1.31 (1.10, 1.58)**	**1.20 (1.10, 1.32)**	**1.17 (1.09, 1.26)**
2	**2.47 (2.02, 3.02)**	**2.51 (2.05, 3.08)**	**1.67 (1.48, 1.88)**	**1.36 (1.23, 1.50)**
≥3	**3.84 (3.23, 4.55)**	**3.01 (2.49, 3.65)**	**2.36 (2.12, 2.64)**	**1.54 (1.40, 1.69)**
Baseline MSK Condition				
Back pain	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)
Knee pain	**1.26 (1.05, 1.52)**	**1.38 (1.14, 1.66)**	**1.26 (1.15, 1.39)**	0.93 (0.86, 1.01)
Hand/wrist pain	0.92 (0.66, 1.28)	1.12 (0.81, 1.54)	1.01 (0.86, 1.19)	**0.78 (0.68, 0.89)**
Inflammatory MSK	**1.73 (1.29, 2.32)**	1.31 (0.91, 1.89)	**1.31 (1.08, 1.59)**	0.9 (0.77, 1.05)
Osteoarthritis	**5.36 (4.23, 6.78)**	**3.69 (2.75, 4.95)**	**2.93 (2.42, 3.55)**	1.16 (0.96, 1.39)
Hip pain	**3.09 (2.38, 4.02)**	**2.02 (1.42, 2.87)**	**1.61 (1.30, 1.98)**	**1.22 (1.03, 1.45)**
Opioids				
No	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)
Yes	**2.72 (2.39, 3.08)**	**2.59 (2.24, 3.00)**	**1.82 (1.68, 1.97)**	**1.35 (1.27, 1.45)**
NSAIDs				

No	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)
Yes	**1.89 (1.66, 2.14)**	**1.66 (1.43, 1.93)**	**1.54 (1.42, 1.67)**	**1.22 (1.15, 1.30)**
Gabapentinoids				
No	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)
Yes	**4.10 (3.25, 5.18)**	**4.45 (3.49, 5.69)**	**2.68 (2.25, 3.19)**	**1.64 (1.39, 1.94)**
Antidepressants				
No	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)
Yes	**2.78 (2.42, 3.20)**	**2.34 (1.98, 2.76)**	**2.09 (1.90, 2.29)**	**1.25 (1.15, 1.36)**
Polypharmacy[e]				
0	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)
1-4	1.20 (1.00, 1.45)	1.03 (0.85, 1.25)	**1.27 (1.14, 1.40)**	**1.09 (1.01, 1.17)**
5-9	**2.16 (1.76, 2.65)**	**1.79 (1.45, 2.22)**	**1.91 (1.71, 2.15)**	**1.19 (1.08, 1.30)**
≥10	**4.06 (3.27, 5.05)**	**3.08 (2.39, 3.96)**	**3.35 (2.88, 3.89)**	**1.27 (1.12, 1.45)**
Smoking Status				
Never	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)
Current	**1.90 (1.63, 2.21)**	**1.42 (1.20, 1.69)**	**1.33 (1.22, 1.44)**	1.06 (0.99, 1.14)
Ex Smoker	**1.51 (1.24, 1.85)**	**1.33 (1.07, 1.65)**	1.01 (0.90, 1.13)	**1.15 (1.06, 1.25)**
Not Recorded	0.94 (0.75, 1.19)	0.86 (0.68, 1.10)	**0.74 (0.66, 0.83)**	0.94 (0.86, 1.02)
BMI[f]				
Underweight/Normal	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)
Overweight	1.21 (1.00, 1.47)	1.10 (0.89, 1.35)	**1.18 (1.06, 1.31)**	**1.17 (1.07, 1.28)**
Obese	**1.58 (1.31, 1.90)**	**1.36 (1.11, 1.67)**	**1.42 (1.28, 1.58)**	**1.13 (1.03, 1.23)**
Not Recorded	0.99 (0.83, 1.20)	0.95 (0.78, 1.15)	**0.89 (0.80, 0.98)**	1.01 (0.93, 1.09)
Modified CCI Score[g]				
0	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)

1	**1.99 (1.71, 2.31)**	**1.33 (1.08, 1.63)**	**1.42 (1.27, 1.58)**	1.03 (0.95, 1.12)
≥2	**2.59 (1.97, 3.40)**	**1.76 (1.27, 2.46)**	**1.78 (1.48, 2.13)**	1.11 (0.94, 1.32)
Contact Count[h]				
1-10	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)
11-15	**1.51 (1.20, 1.90)**	1.09 (0.85, 1.38)	**1.36 (1.20, 1.54)**	**1.18 (1.08, 1.28)**
16-25	**1.76 (1.44, 2.15)**	1.20 (0.97, 1.47)	**1.72 (1.55, 1.92)**	**1.19 (1.10, 1.29)**
≥26	**3.09 (2.60, 3.66)**	**2.31 (1.93, 2.77)**	**2.60 (2.34, 2.87)**	**1.29 (1.20, 1.39)**

Abbreviations: IMD = Index of Multiple Deprivation; MSK = Musculoskeletal; NSAIDs = Non-Steroidal Anti-Inflammatory Drugs; BMI = Body Mass Index; CCI = Charlson Comorbidity Index

Notes: Values are presented as unadjusted odds ratios with 95% confidence intervals.

Statistically significant estimates (where 95% CI doesn't include the value 1) are shown in bold

[a] All odds ratios are calculated with respect to the reference trajectory: Single.

[b] North of England defined as: Northeast, Northwest, Yorkshire and the Humber; Middle of England: East Midlands, West Midlands, East of England; and South of England: Southeast, Southwest, London

[c] Quintiles are used for IMD (1-5), where a higher score represents more deprived areas

[d] Excluding the index MSK consultation

[e] Excluding Opioids, NSAIDs, Gabapentinoids, Antidepressants

[f] Underweight/Normal: 10<=BMI<25; Overweight: 25<=BMI<30; Obese: 30<=BMI<80; Not Recorded: no BMI data available or BMI<10 or BMI>=80

[g] Excluding Rheumatic Disease

[h] Count of all medical consultations (except MSK related), as well as any recording of data such as BMI, smoking etc.

Table N.2. Characteristics Associated with Optimal Trajectories of Work Absence Due to a MH Condition Using the 'Single' Trajectory Class as the Reference (Unadjusted Model)

	Trajectory Class			
	Chronic Sustained[a] n=2,881	**Chronic Fast Decreasing[a]** n=3,848	**Intermittent Low[a]** n=10,835	**Short Term[a]** n=21,534
Sex				
Male	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)
Female	**0.75 (0.69, 0.82)**	**0.81 (0.74, 0.88)**	**1.08 (1.02, 1.15)**	**0.93 (0.89, 0.97)**
Age				
16-25 years	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)
26-35 years	**0.77 (0.68, 0.88)**	0.96 (0.84, 1.09)	**0.83 (0.77, 0.90)**	**1.33 (1.25, 1.42)**
36-45 years	0.99 (0.87, 1.13)	**1.26 (1.10, 1.43)**	**0.75 (0.68, 0.82)**	**1.53 (1.43, 1.64)**
46-55 years	**1.33 (1.16, 1.51)**	**1.70 (1.49, 1.93)**	**0.83 (0.76, 0.92)**	**1.68 (1.56, 1.81)**
56-66 years	**1.88 (1.59, 2.22)**	**2.10 (1.79, 2.47)**	1.02 (0.89, 1.16)	**1.74 (1.57, 1.92)**
Region[b]				
South of England	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)
North of England	**1.56 (1.37, 1.77)**	**1.81 (1.63, 2.01)**	**1.12 (1.04, 1.22)**	**1.27 (1.19, 1.35)**
Middle of England	**1.58 (1.39, 1.80)**	**1.37 (1.22, 1.54)**	0.99 (0.91, 1.08)	**1.18 (1.11, 1.25)**
IMD[c]				
1	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)
2	**1.25 (1.03, 1.50)**	**1.22 (1.04, 1.43)**	**1.15 (1.03, 1.29)**	0.96 (0.89, 1.04)
3	**1.56 (1.30, 1.87)**	**1.40 (1.20, 1.64)**	**1.35 (1.21, 1.51)**	0.94 (0.87, 1.01)
4	**1.96 (1.64, 2.35)**	**1.57 (1.33, 1.84)**	**1.66 (1.49, 1.85)**	0.94 (0.87, 1.01)
5	**3.10 (2.63, 3.66)**	**2.35 (2.03, 2.73)**	**2.06 (1.86, 2.29)**	0.95 (0.88, 1.02)

Missing	0.72 (0.52, 1.00)	0.94 (0.73, 1.22)	**0.63 (0.49, 0.81)**	1.08 (0.97, 1.21)
MH Consultations - Prior 2 Years[d]				
0	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)
1	**1.22 (1.08, 1.37)**	**1.15 (1.03, 1.29)**	**1.28 (1.18, 1.39)**	**0.93 (0.87, 0.99)**
2	**1.48 (1.26, 1.72)**	**1.28 (1.09, 1.49)**	**1.43 (1.28, 1.59)**	0.92 (0.84, 1.01)
≥3	**1.60 (1.41, 1.81)**	**1.23 (1.08, 1.40)**	**1.77 (1.63, 1.93)**	**0.87 (0.81, 0.94)**
Baseline MH Condition				
Stress	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)
Anxiety and Depression	**4.03 (3.47, 4.68)**	**3.15 (2.77, 3.59)**	**2.09 (1.90, 2.29)**	**1.39 (1.30, 1.48)**
Depression	**3.82 (3.30, 4.43)**	**2.75 (2.42, 3.13)**	**1.93 (1.77, 2.11)**	**1.29 (1.21, 1.38)**
Anxiety	**1.95 (1.67, 2.27)**	**1.40 (1.22, 1.60)**	**1.33 (1.22, 1.46)**	0.98 (0.93, 1.05)
Opioids				
No	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)
Yes	**1.93 (1.72, 2.16)**	**1.52 (1.34, 1.72)**	**1.78 (1.63, 1.94)**	1.00 (0.93, 1.08)
NSAIDs				
No	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)
Yes	**1.29 (1.14, 1.47)**	1.11 (0.98, 1.25)	**1.30 (1.19, 1.41)**	1.02 (0.96, 1.10)
Gabapentinoids				
No	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)
Yes	**1.97 (1.51, 2.59)**	**1.48 (1.08, 2.03)**	**1.94 (1.58, 2.38)**	0.89 (0.73, 1.08)
Antidepressants				
No	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)
Yes	**1.66 (1.52, 1.81)**	**1.34 (1.22, 1.46)**	**1.70 (1.59, 1.81)**	**0.91 (0.87, 0.96)**
Polypharmacy[e]				

0	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)
1-4	**0.86 (0.77, 0.97)**	**0.84 (0.76, 0.94)**	1.09 (1.00, 1.19)	0.96 (0.91, 1.02)
5-9	1.06 (0.93, 1.21)	0.93 (0.82, 1.06)	**1.45 (1.31, 1.59)**	0.93 (0.87, 1.00)
≥10	**1.74 (1.46, 2.08)**	**1.41 (1.19, 1.67)**	**2.31 (2.03, 2.64)**	**0.89 (0.80, 0.99)**
Smoking Status				
Never	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)
Current	**2.17 (1.96, 2.41)**	**1.87 (1.69, 2.07)**	**1.57 (1.46, 1.69)**	**1.09 (1.03, 1.15)**
Ex Smoker	1.16 (0.99, 1.36)	**1.18 (1.03, 1.36)**	1.07 (0.97, 1.19)	**1.15 (1.07, 1.24)**
Not Recorded	**1.26 (1.09, 1.46)**	**1.24 (1.08, 1.41)**	1.07 (0.98, 1.18)	**1.08 (1.01, 1.16)**
BMI[f]				
Underweight/Normal	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)
Overweight	1.10 (0.96, 1.27)	1.09 (0.95, 1.24)	0.96 (0.88, 1.05)	**1.13 (1.06, 1.20)**
Obese	**1.61 (1.41, 1.85)**	**1.47 (1.29, 1.68)**	**1.32 (1.20, 1.45)**	**1.21 (1.13, 1.31)**
Not Recorded	**1.26 (1.12, 1.42)**	**1.26 (1.13, 1.40)**	0.99 (0.91, 1.06)	1.06 (1.00, 1.12)
CCI Score				
0	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)
1	**1.39 (1.23, 1.57)**	1.11 (0.97, 1.26)	**1.45 (1.33, 1.59)**	0.99 (0.92, 1.06)
≥2	**1.67 (1.31, 2.12)**	**1.42 (1.11, 1.82)**	**1.71 (1.43, 2.05)**	1.08 (0.93, 1.26)
Contact Count[g]				
1-10	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)
11-15	1.02 (0.87, 1.19)	1.06 (0.93, 1.20)	**1.18 (1.07, 1.31)**	1.05 (0.98, 1.12)
16-25	1.14 (1.00, 1.30)	1.00 (0.89, 1.13)	**1.43 (1.30, 1.57)**	0.97 (0.91, 1.04)
≥26	**1.51 (1.34, 1.69)**	**1.22 (1.09, 1.37)**	**1.98 (1.82, 2.16)**	0.96 (0.91, 1.02)

Abbreviations: IMD = Index of Multiple Deprivation; MH = Mental Health; NSAIDs = Non-Steroidal Anti-Inflammatory Drugs; BMI = Body Mass Index; CCI = Charlson Comorbidity Index

Notes: Values are presented as unadjusted odds ratios with 95% confidence interval.

Statistically significant estimates (where 95% CI doesn't include the value 1) are shown in bold

[a] All odds ratios are calculated with respect to the reference trajectory: Single.
[b] North of England defined as: Northeast, Northwest, Yorkshire and the Humber; Middle of England: East Midlands, West Midlands, East of England; and South of England: Southeast, Southwest, London
[c] Quintiles are used for IMD (1-5), where a higher score represents more deprived areas
[d] Excluding the index MH consultation
[e] Excluding Opioids, NSAIDs, Gabapentinoids, Antidepressants
[f] Underweight/Normal: 10<=BMI<25; Overweight: 25<=BMI<30; Obese: 30<=BMI<80; Not Recorded: no BMI data available or BMI<10 or BMI>=80
[g] Count of all medical consultations (except MH related), as well as any recording of data such as BMI, smoking etc.