TITLE

Predicting negative health outcomes in older general practice patients with chronic illness: rationale and development of the PROPERmed harmonized individual participant data database

AUTHOR NAMES AND AFFILIATIONS

*Ana I. González-González[1,2], *Truc S. Nguyen[1], Andreas D. Meid[3], Jeanet W. Blom[4], Marjan van den Akker[1,5,15], Petra J.M. Elders[6], Ulrich Thiem[7], Daniela Kuellenberg de Gaudry[8], Kym Snell[10], Rafael Perera[11], Karin M.A. Swart[6], Henrik Rudolf[12], Donna Bosch-Lenders[5], Hans-Joachim Trampisch[12], Joerg J. Meerpohl[8,9], Benno Flaig[1], Ghainsom Kom[13], Ferdinand M. Gerlach[1], Walter E. Hafaeli[3], Paul P. Glasziou[14], Christiane Muth[1]

[1]Institute of General Practice, Goethe University Frankfurt, 60590 Frankfurt am Main, Germany

[2]Red de Investigación en Servicios de Salud en Enfermedades Crónicas (REDISSEC), Madrid, Spain

[3]Department of Clinical Pharmacology and Pharmacoepidemiology, University Hospital Heidelberg, 69120 Heidelberg, Germany

[4]Department of Public Health and Primary Care, Leiden University Medical Center, 2300RC Leiden, The Netherlands

[5]School of CAPHRI, Department of Family Medicine, Maastricht University, 6211 LK Maastricht, The Netherlands

[6]Department of General Practice and Elderly Care Medicine, Amsterdam University Medical Center, 1007 MB Amsterdam, The Netherlands

[7]Chair of Geriatrics and Gerontology, University Clinic Eppendorf, 20246 Hamburg, Germany

[8]Institute for Evidence in Medicine (for Cochrane Germany Foundation), Medical Center – University of Freiburg, 79110 Freiburg, Germany

[9]Cochrane Germany, Cochrane Germany Foundation, Breisacher Strasse 153, 79110 Freiburg, Germany

[10]Centre for Prognosis Research, Research Institute for Primary Care & Health Services, Keele University, Staffordshire ST5 5BG, United Kingdom

[11]Nuffield Department of Primary Care, University of Oxford, Oxford OX2 6GG, United Kingdom

[12]AMIB, Ruhr University, 44780 Bochum, Germany

[13]Techniker Krankenkasse (TK), 22765 Hamburg, Germany

[14]Centre for Prognosis Research in Evidence-Based Practice, Bond University, Robina QLD 4226, Australia

[15]Academic Centre for General Practice, Department of Public Health and Primary Care, KU Leuven, Belgium

*both authors contributed equally

CORRESPONDING AUTHOR

Ana I. González-González.

Institute of General Practice, Goethe University Frankfurt, 60590 Frankfurt am Main, Germany

Email: gonzalezgonzalez@allgemeinmedizin.uni-frankfurt.de

ABSTRACT

The prevalence of multimorbidity and polypharmacy increases significantly with age, and are associated with negative health consequences. However, most current interventions to optimize medication have failed to show significant effects on patient-relevant outcomes. This may be due to ineffectiveness of interventions themselves, but may also reflect other factors: insufficient sample sizes, heterogeneity of population. To address this issue, the international PROPERmed collaboration was set up to obtain/synthesize individual participant data (IPD) from five cluster-randomized trials. The trials took place in Germany and The Netherlands and aimed to optimize medication in older general practice patients with chronic illness. PROPERmed is the first database of IPD to be drawn from multiple trials in this patient population and setting. It offers the opportunity to derive prognostic models with increased statistical power for prediction of patient-relevant outcomes resulting from the interplay of multimorbidity and polypharmacy. This may help patients from this heterogeneous group to be stratified according to risk and enable clinicians to identify patients that are likely to benefit most from resource/time-intensive interventions. The aim of this manuscript is to describe the rationale behind PROPERmed collaboration, characteristics of the included studies/participants, development of the harmonized IPD database and challenges faced during this process.

Registration

PROSPERO ID: CRD42018088129

Keywords (max. 6)

Elderly, hospitalization, meta-analysis, multimorbidity, polypharmacy, prognosis, quality of life.

INTRODUCTION

Older people with multiple chronic diseases (multimorbidity) are the norm rather than the exception, and up to 80% of consultations in general practice are devoted to helping them (1, 2). These patients are frequently treated with multiple drugs (polypharmacy), which increases the risk of both inappropriate prescriptions and medication underuse (3-7). Both multimorbidity and polypharmacy are associated with a range of negative health outcomes such as falls, hospital admissions, reduced health-related quality of life (HRQoL), decreased functionality and loss of autonomy (8-13). Multimorbidity and polypharmacy have been recognized as major challenges facing health care systems, as they not only worsen health outcomes in individuals, but lead to higher costs, e.g., for health care utilization, treatment of adverse drug reactions (ADR), and home care (13-15).

To optimize the care and treatment of chronically ill patients, various interventions have been developed and tested, of which most were complex and resource intensive. However, most cluster-randomized controlled trials (cRCT) evaluating complex interventions associated with polypharmacy and multimorbidity in general practice have failed to demonstrate significant effects on patient-relevant outcomes (16). While positive effects have been shown on process parameters (e.g., outcomes related to prescribing quality), interventions have showed few and inconsistent effects on clinical outcomes (e.g., falls), and no effect on well-being (e.g., quality of life) (17-21). Several factors may be responsible for this. Firstly, the intervention may have been ineffective or of inadequate intensity, or its implementation may have failed, perhaps because the follow-up period was too short to allow for behavioural change in the prescribers. Secondly, the study design and the methodology may have been flawed, possibly because cross-sectional data on frequently changing medication regimens was not collected for long enough, or due to contamination. Thirdly, the Hawthorne effect may have played a role, and the simple fact of participation in a trial may have resulted in an improvement in patients' health (19-24). Furthermore, existing definitions of multimorbidity and polypharmacy lead to heterogeneity in this patient group because they cover a broad range of patients,

ranging from the fit and active that are in a good state of health despite their diseases and prescriptions, to those with a high disease and treatment burden and negative outcomes. Heterogeneity in this patient population makes it particularly difficult to target the population that may benefit most from interventions that are time- and resource-intensive and often complex (25).

The use of prognostic models can help predict the risk of specific future outcomes in individual patients and may therefore help stratify risk and identify patients in need (26). A systematic review of existing tools to predict risk found models aimed at identifying the risk of adverse events in patients with multimorbidity, but none of them adequately considered polypharmacy-related risks and the interrelation of multiple conditions and medication (27-30). In general, prognostic models that have taken multimorbidity into account have shown only modest predictive accuracy. The authors of the review identified three main limitations to implementing prognostic models in this setting - ethical and data privacy issues, a lack of acceptance for such models among health professionals, and a lack of evidence of their (cost-) effectiveness (31). Using risk assessment tools is felt to be beneficial for initiating discussion, engaging patients in risk discussions, and guiding both health professionals and patients around decision-making (32). However, the little effect has been found on patient attitudes who express difficulties trying to understand the information provided with this type of tools (33).

To support the creation of prognostic models that address the problems mentioned above, a group of international researchers that are active in the field of multimorbidity and polypharmacy in general practice set up the PROPERmed collaboration and constructed the first major database of individual patient information that was drawn from multiple trials in this population. As the approach offers both statistical and clinical advantages, our aim was to provide a harmonized individual patient database to support the development and validation of prognostic models, (34). In this manuscript we describe the rationale behind the PROPERmed collaboration, the characteristics of the included studies and patients, and the design, development and challenges of creating a harmonized individual participant data (IPD) database.

THE PROPERMED COLLABORATION

The PROPERmed collaboration included trialists from five study centers in Germany and The Netherlands that shared strong links due to pursuing a common interest in multimorbidity and polypharmacy research: i) Goethe University Frankfurt/Main (Germany), ii) Amsterdam University Medical Center (The Netherlands), iii) Maastricht University (The Netherlands), iv) Leiden University Medical Center (The Netherlands), and v) Ruhr University Bochum (Germany). A collaborative approach was used in preference to a systematic approach (35) because this was the first time a meta-analysis of data from this population had been conducted, and the ability to confront unknown difficulties required the expertise of a group of researchers with an established common interest. All the trials had been registered and approved by the respective Medical Ethics Committees. The PROPERmed collaboration further involved German experts from Heidelberg University Hospital and the University of Freiburg (Cochrane Germany), as well as an international scientific advisory board that included experts from Bond University, Gold Coast (Australia), as well as the University of Oxford and the Centre for Prognosis Research, Keele University (UK). Key barriers of successfully implementing prognostic models not only include a lack of acceptance among health professionals (31), but also a lack of support by leading stakeholders in the field of application as well as patients' acceptance . Furthermore, a timely application of prognostic models in routine care seems to facilitate implementation (26). Therefore, to ensure the stakeholder perspective was not neglected, the Techniker Krankenkasse (TK) statutory health insurance fund was also involved in the collaboration. By collaborating with the TK, we will be able to gain relevant insights with regard to the potential implementation of our model considering the macro, meso and micro level: i) Macro: German statutory health insurances are represented in the Federal Joint Committee (G-BA). The G-BA consists of umbrella organizations of the self-governing healthcare system including patient representatives and is responsible for the definition of healthcare services to be paid by statutory health insurances,

and quality assurance of care; ii) Meso: Depending on the model, the included variables and defined outcomes, the potential application of the prognostic model to data from routine care provided by statutory health insurances will be discussed; iii) Micro: As not all variables included in the final model may be available in routine data from statutory health insurances, additional possibilities of data collection including patients will be discussed (e.g., via patient apps) in order to complement available data from routine care.

SOURCE OF DATA – CHARACTERISTICS OF THE PROPERmed STUDIES

The PROPERmed study included IPD from five cRCTs: PRIMUM (24), Opti-Med (36, 37), PIL (Nederlands Trial Register, NTR2154) (38), ISCOPE (39) and RIME (Deutsches Register Klinischer Studien-ID, DRKS00003610). The cRCTs were conducted from 2009 onwards in general practices from The Netherlands (ISCOPE, Opti-Med, and PIL) and Germany (PRIMUM and RIME) and evaluated complex interventions aimed at optimizing (pharmacological) treatment in older patients with chronic disease. Practices served as the unit of randomization in all the trials. Four of the five cRCTs used a pragmatic parallel-group design and one used a stepped-wedge design (PIL). All the trials compared a structured medication review (complex intervention) with usual care except one (ISCOPE), which compared usual care with an integrated care plan that used a functional geriatric approach. Table 1 provides an overview of the characteristics of the included trials.

[About here: Table 1 on Main characteristics of the included trials].

DEFINITION OF THE CORE DATA SET

At a kick-off meeting in May 2017, we and the other collaborators decided upon the data to be collected for the IPD-MA based on the availability of relevant data in the trials. We agreed upon overall inclusion and exclusion criteria for practices and patients, as well as a core data set covering a variety

of candidate prognostic factors and patient-relevant outcomes to be predicted in the subsequent modelling approach.

Participants

Practices

To be included in the trials, general practices had to provide primary care within the framework of the German or Dutch statutory health insurance systems. In one study (PRIMUM), practices specializing in unconventional treatments or special conditions (e.g., HIV) were excluded. No additional exclusion criteria were applied to practices in the other trials.

Patients

Based on the inclusion criteria in the trials, we included patients in PROPERmed that were aged 60 years and older, had been diagnosed with at least one chronic condition, and had at least one chronic prescription at study baseline. We defined chronic conditions in accordance with O´Halloran et al. (40), and defined chronic prescriptions in the same way as the trials (two weeks in PRIMUM, two months in ISCOPE, and three months in Opti-Med, PIL and RIME).

In accordance with the exclusion criteria used in the trials, critically ill patients with a limited life expectancy were not included. Additional inclusion and exclusion criteria were specified in individual trials, such as abuse of alcohol or illegal drugs, which was an exclusion criterion in PRIMUM and PIL. Between-trial heterogeneity resulting from the impact of trial-specific criteria was considered to be acceptable.

Outcomes and candidate prognostic variables

For the planned prediction models, we agreed upon outcomes signifying deterioration in health-related quality of life (dHRQoL) and all-cause hospital admissions at six-month follow up. Following a discussion with the co-authors, the outcomes of interest were selected based on their relevance and availability across the included trials. We also selected a full range of candidate prognostic variables for the models, including data on socio-demographics and lifestyle, patient morbidity, medication,

functional status and well-being (HRQoL, physical and cognitive components), and whether patients had been hospitalized in the 12 months previous to study enrolment.

DATA COLLECTION AND HARMONIZATION

Regardless of format, IPD for all five included cRCTs, the lists of variables, and information on data collection, were gathered and stored in a secure data repository. As the measuring instruments used in the studies varied, ex-post data harmonization procedures were conducted before model development. As data harmonization aimed to preserve as much information as possible, missing data, inconsistencies, and errors that occurred during data preparation were queried and discussed with the original authors. Variables from the original trials were also harmonized using different techniques, which included the recoding of categories and labels, and the clinical harmonization of variables (e.g., see 'morbidity-related candidate prognostic variables'). Whenever necessary, new variables were created from other available information by, for example, using cut-offs (see 'depressive symptoms'). If specific trials used psychometric instruments based on differing constructs to measure the same outcome (see 'functional status and well-being'), the underlying psychometrical constructs were preserved and standardized mean differences used to render them comparable. Table 2 provides more detailed information on the data harmonization process.

[About here: Table 2 on Data harmonization].

Health-related quality of life (HRQoL)

At baseline and after the six-month follow-up, all trials used the EQ-5D-3L index score (41) to measure generic HRQoL in the dimensions mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. In two of the Dutch trials (PIL and ISCOPE), the translated question on "mobility" had to undergo slight linguistic adaptation because the original item in the standardized questionnaire frequently resulted in missing values in older populations (42). We calculated EQ-5D-3L index scores based on time trade-off norm values (43) and defined the binary outcome dHRQoL as a decrease in

213 the EQ-5D-3L index score of at least five percent between baseline and six-month follow-up. We
214 considered this cut-off to be clinically relevant because it was derived from observations of the minimal
215 important difference (MID) in HRQoL deterioration that is perceived by patients. This is a rather
216 conservative assumption, as MIDs in the populations under investigation varied depending on disease
217 severity and age (44, 45).

218 **All-cause hospital admissions**

219 We defined hospital admission as the moment when a patient receives inpatient services in a hospital
220 for any reason between baseline and six-month follow-up. As some trials provided no precise
221 information on reasons for admission and length of stay, hospitalizations included emergency
222 admissions and planned admissions from waiting lists, both day and night. Trials either used patient-
223 reported data from questionnaires (Opti-Med, PIL, ISCOPE and RIME) or physician-reported case report
224 forms (PRIMUM) to collect information on hospital admissions. All trials except Opti-Med collected
225 baseline information on previous hospitalizations, with PRIMUM, ISCOPE and RIME providing
226 information on the number of admissions in the twelve-month period before baseline. With respect
227 to follow-up information, admissions between baseline and six-month follow-up were recorded in
228 PRIMUM, Opti-Med and RIME, and between baseline and 12-month follow-up in PIL and ISCOPE. The
229 number of admissions at baseline was considered to be a continuous variable, while the number of
230 follow-up admissions was operationalized as a binary outcome, as it resulted in fewer missing values.

231 **Sociodemographic and lifestyle-related variables**

232 All trials collected sociodemographic and lifestyle-related variables at baseline using self-administered
233 questionnaires (PRIMUM, Opti-Med, PIL and ISCOPE), or patient interviews (RIME). Trials recorded the
234 age of the patient when recruited, and gender. We categorized living situation as “living at home” or
235 “institutionalized living”. We categorized educational level according to the aggregated levels of the
236 International Standard Classification of Education (ISCED) issued by the United Nations Educational,
237 Scientific and Cultural Organization (UNESCO) (46). The UNESCO recognizes seven levels of education

238 in ISCED, from Level 0 (pre-primary education) to Level 6 (second stage of tertiary education), which
239 can be collapsed into three aggregated levels: low, medium and high. Three trials (PRIMUM, PIL, and
240 RIME) gathered information on the smoking status of their patients. We classified smoking status as
241 current smoker, ex-smoker, or non-smoker.

242 **Morbidity-related candidate prognostic variables**

243 Information on diagnoses was collected at baseline in all trials except PIL, which reported it at the start
244 of the intervention. Three studies collected morbidity data using international classifications: Two
245 studies provided GP-reported data (PRIMUM using ICD-10 codes (47) and PIL using ICPC-1 codes (48)).
246 ISCOPE took patients' chronic diagnoses from electronic medical records (EMR), defined them
247 according to O'Halloran et al. (40), and coded them based on the second version of the International
248 Classification of Primary Care (ICPC-2) (49). The remaining two trials used a limited list of conditions,
249 which were either taken from EMR (Opti-Med, 26 conditions) or during patient interviews (RIME, 28
250 conditions). In PROPERmed, we clinically harmonized the data on medication and used the ICPC-2 to
251 describe chronic conditions: ICPC-1 (PIL) and ICD-10 (PRIMUM) codes were thus converted into the
252 corresponding ICPC-2 codes, and the respective codes were assigned to the list of non-coded chronic
253 conditions described in Opti-Med and RIME.

254 Multimorbidity can be expressed in terms of a disease count, based on multimorbidity scores or
255 multidimensional assessment methods. Before using these instruments in PROPERmed, they had to be
256 modified according to the common chronic conditions that were assessed in all the trials. See Table 3
257 for more detailed information on proxies for multimorbidity instruments.

258 [About here: Table 3 on Proxies for multimorbidity instruments]

259 *Modified Diederichs disease count*

260 The Diederichs list (50) contains 17 chronic conditions, which, according to a systematic review, are
261 considered in 39 different multimorbidity indexes. Our modified Diederichs list contains 12 (71 %) of
262 the 17 conditions in the original instrument.

Modified Physical Component Summary score (PCS) and Mental Component Summary score (MCS)

The recently published Health-related Quality of life Comorbidity Index (HRQoL-CI) consists of two independent subscales (51-53). The comorbidity subscale for physical HRQoL (PCS) considers 20 groups of chronic conditions, while the subscale for mental HRQoL (MCS) consists of 15 groups of conditions. We derived modified PCS and MCS scores (51, 52) depending on the chronic conditions considered in all five cRCTs. The modified PCS score included 12, and the MCS 6 groups of chronic conditions (60 % and 40 % respectively of the total count in the original instrument).

Modified Charlson Comorbidity Index

The Charlson Comorbidity Index (CCI) (54) is the most commonly used scale to measure multimorbidity in hospital admission predicting models (31). The CCI assesses the comorbidity level by considering both the number and severity of 19 pre-defined comorbid conditions. It provides a weighted score of a patient's morbidities, which can be used to predict such short- and long-term outcomes as functional status, length of stay in hospital, and mortality rates. In general, 0-1 points indicates the absence of any comorbidity, 2 points indicates mild comorbidity and a score of ≥ 3 indicates severe comorbidity. We considered the sum-scores of the CCI provided in two of the trials (PRIMUM and RIME). The CCI modified for use in the PROPERmed predicting models did not include dementia, as only one trial (ISCOPE) included patients with this diagnosis.

Medication-related candidate prognostic variables

All trials obtained medication data as patient-reported data at baseline except PIL and ISCOPE, in which the general practitioner reported the medications at baseline (ISCOPE), or at the start of the intervention (PIL).

Potentially inappropriate medications (PIM) and potential prescribing omissions (PPOs)

Medication inappropriateness and omissions were assessed using several instruments: the EU-PIM list (55), STOPP-START criteria (56), the Dreischulte high-risk prescribing list (57), the Anticholinergic Drug Scale (ADS) (58, 59), the Drug Burden Index (DBI) (60-62) and the Anticholinergic Drug Burden

288 according to Duran et al (63). From those lists, we measured: 1) any PIM or PPO (yes/no); 2) number
289 of PIMs or PPOs prescribed per patient; 3) individual scores measured by the instruments, when
290 applicable. Instruments included in PROPERmed were modified, as missing information in (some of)
291 the trials, such as daily doses, duration of use, and information on severity of disease, meant they did
292 not represent the full range of PIMs and PPOs considered in the original instruments.

293 **Functional status and well-being**

294 *Functional status*

295 All trials used patient questionnaires to obtain patient-reported data on functional status at baseline.
296 Two trials (ISCOPE and PIL) measured functional status using the Katz-15 (combination of KATZ-6 and
297 Lawton IADL) questionnaire (64), two trials (PRIMUM and RIME) used the 13-item vulnerable elderly
298 survey (VES-13) (65) and one trial (Opti-Med) measured functional status using the Geriatric Giants
299 VAS (GGV) scale (0-10) (66) for mobility problems, which was created *ad hoc*. We standardized the
300 original instruments' scales statistically, thus enabling the information to be included in analysis
301 models, irrespective of differences in the underlying constructs of the instruments.

302 *Pain*

303 Using different approaches, pain was assessed at baseline in all trials. One trial (PRIMUM) measured
304 pain using the von Korff index (67), one trial (Opti-Med) used a VAS scale that had been elaborated *ad*
305 *hoc* (from 0 to 10), another (ISCOPE) used a single question (pain yes or no), two trials (Opti-Med and
306 RIME) used the pain-related question from SF-12 (68, 69) and one (PIL) used the pain-related question
307 from SF-36 (70). We constructed categorical variables for pain by clinically harmonizing the different
308 instruments used across trials as follows: a) Pain intensity was classified as "no pain", "low intensity
309 pain" and "high intensity pain", based on data from PRIMUM (67), Opti-Med and PIL (68); b) Pain
310 disability was categorized as "no disability from pain", "disability of low intensity" and "disability of
311 high intensity" based on data from PRIMUM (67), Opti-Med, RIME (68) and PIL (70); c) Binary
312 classification of pain (i.e. pain vs. no pain) was based on data from PRIMUM (67), Opti-Med, PIL,

ISCOPE and RIME (68); and d) Pain according to SF surveys was based on data from Opti-Med, RIME (68) and PIL (70) and did not require harmonization.

Depressive symptoms

All trials used several questionnaires at baseline to assess depressive symptoms. Three trials used the Geriatric Depression Scale (GDS) to assess symptoms of depression (PRIMUM and ISCOPE used the GDS with 15 items (71), and RIME used it with 5 items (72)). Two trials (Opti-Med and RIME) used the depressive symptoms-related question on SF-12 (68), and PIL used the depressive symptoms-related question on SF-36 (70). In addition to standardized forms of the various instruments, we also derived categorical variables by considering the cut-offs of the original instruments as follows: a) A binary classification of depressive symptoms (i.e. depressive symptoms vs. no depressive symptoms) was carried out using data from PRIMUM and ISCOPE (GDS-15 cut-off of ≥ 5 for depressive symptoms) (71, 73) and RIME (GDS-5 cut-off of ≥ 2 for depressive symptoms) (72); b) The binary classification of depressive symptoms was also performed for Opti-Med and RIME (SF-12 cut-off of < 42 for depressive symptoms) (74) and for PIL (SF-36 cut-off of < 50.2 for depressive symptoms) (75).

RISK OF BIAS ASSESSMENT

A risk of bias assessment (RoB) was conducted for each included trial in order to assess possible differences in the definitions and measuring methods of outcomes and variables across all five cRCTs. The RoB was determined using PROBAST (Prediction model Risk of Bias Assessment Tool), which is a qualitative assessment tool developed for evaluating both RoB and the applicability of prognostic prediction models (76). It includes the following domains: 1. participant selection, 2. predictors, 3. outcomes, 4. sample size and participant flow, and 5. analysis. Since not all domains were relevant to the included cRCTs, the robustness of evidence was assessed for the relevant ones by one researcher (DKDG) and checked by another (AIGG). Any inconsistency was resolved by discussion.

Based on PROBAST, the overall RoB assessment of the included trials was “unclear”, mainly due to uncertainty about the definition and assessment method of possible predictor variables reported in the studies. However, differences in definitions and measurement methods were also considered and accounted for during the data harmonization process. Since the PROPERmed dataset was designed on the basis of the available data from the trials, the applicability of the data from the cRCTs in the prognostic models was automatically rated as of “low concern”.

See Table 4 for more detailed information about the RoB assessment.

[About here: Table 4 on Risk of bias assessment with the PROBAST tool].

CHARACTERISTICS OF THE PROPERMED STUDY POPULATION

In total, 6,139 patients were potentially eligible for inclusion in the PROPERmed IPD-MA dataset. Two hundred and eighteen patients did not meet the inclusion criteria and 1,360 patients did not provide relevant data at baseline. After excluding these patients from the IPD-MA, we finally included 4,561 patients from 307 general practices. See Figure 1 for more detailed information about the selection and maintenance of PROPERmed participants.

[About here: Figure 1 on Selection and maintenance of PROPERmed participants]

The median age of the PROPERmed population was 78 years and ranged from 60 to 102 years. Women represented 58% of patients in the IPD-MA. Most participants (95%) lived at home, had a low or medium educational level (89%) and had smoked (49%), or were current smokers (43%).

The median number of chronic diseases according to the modified Diederichs list (50) was three and ranged from zero to ten chronic diseases per patient. The most prevalent chronic conditions were hypertension (67%), diabetes (36%), coronary heart disease (34%), osteoarthritis (34%), and vision problems (32%).

The median number of drugs per patient was seven and ranged from one to 28. According to the EU-PIM list (55), 2,624 (58%) of 4,561 patients were taking at least one inappropriate medication, with a

median of one per patient and ranging from zero to seven. The number of patients with at least one PIM according to the STOPP criteria (56), and one PPO according to START criteria (56) was 3,899 (86%) and 2,634 (58%) patients respectively. Regarding anticholinergic medication, 1,311 (29%), 1,646 (36%) and 2,269 (50%) patients were respectively taking at least one drug included on the ADB (63), DBI (60-62) and ADS score (58, 59) lists. The most prevalent ATC codes (3-digit level) representing chronic prescriptions were B01 – Antithrombotic agents (64%), C09 - Agents acting on the renin-angiotensin system (63%), C10 - Lipid modifying agents (53%), C07 – Beta blocking agents (53%) and C03 – Diuretics (41%).

Of the included patients, 68% experienced some pain, and 15% / 21% had symptoms suggesting depression according to GDS (71) / SF-12 (68, 69) and SF-36 (70), respectively.

Table 5 provides a detailed description of patient characteristics based on variables related to candidate prognostic variables and outcomes.

The mean EQ-5D-3L index at baseline and six-month follow-up was 0.73 (0.25) in the PROPERmed population. The number of patients that had been admitted to hospital at least once during the six-month follow-up period was 404 of the 3,296 patients (21%) that participated in PRIMUM, ISCOPE and RIME.

[About here: Table 5 on Descriptives of outcomes and candidate prognostic variables].

DISCUSSION

PROPERmed is the first study to combine the data of older patients with multiple conditions and chronic medication prescriptions from five cRCTs. Increasing case-mix variability is generally considered to raise the performance of prognostic models. However, systematic between-study heterogeneity resulting from different time frames and health care policies may be a problem. Thus, it is certainly a major advantage that the PROPERmed dataset includes trials completed after the year 2009 in The Netherlands and Germany. The reduction to two countries of origin within a reasonable

time frame reduces variation between drug markets and ensures that the prescribing preferences and patterns of GPs vary little and are at low risk of being biased by, for example, differences in the availability of medicines (77-80). On the other hand, the generalizability of the results from the PROPERmed prognostic models may be affected by similarities in the inclusion and exclusion criteria across the cRCTs. External validation in another population from another setting or in another timeframe is therefore important.

In IPD-based research, data harmonization is absolutely essential to subsequent analyses. This required enormous effort but resulted in comprehensive data, which is essential if data is to be used in prognostic models. Furthermore, data harmonization minimizes the potential RoB of individual studies. In order to generate comparable morbidity counts or indexes, PIM or PPO lists, and scores across all the trials, we had to modify some of the validated instruments. In this particular example, limited information on morbidity (Opti-Med and RIME) and medication (doses, duration of use, and severity of diseases) were significant obstacles and had to be overcome. We met the challenge of the use of different scales to measure functional status and depressive symptoms by standardizing the original instruments. The use of standardized measures means it will be slightly more difficult to interpret the results. Furthermore, the measures will have to be back translated to the original instruments for straightforward clinical interpretation.

In projecting the future success of any analysis, a lack of information on specific reasons for hospitalization may well be significant. Unfortunately, the available information prevents us from predicting unplanned hospitalizations, although this is highly relevant and would lead to better performing models. Instead we will only be able to predict emergency admissions and planned admissions simultaneously, and will not be able to distinguish between day- and night-time admissions. This may be seen as a limitation to the further use of our model.

Other research collaborations have been confronted with the challenges of data harmonization. The European Project on Osteoarthritis (EPOSA) (81), which focuses on personal and societal burden and

412 the determinants of osteoarthritis, merged the data from cohort studies in five European countries.
413 They described the entire harmonization processes in a harmonization guideline and highlighted the
414 difficulties that arose from differences in wording and categories, and in the classifications and absence
415 of data. They also stressed how important it is to make the original data accessible to ensure the ex-
416 post harmonized data is easy to interpret (81). As in PROPERmed, the design of the International
417 Mission for Prognosis and Clinical Trial (IMPACT) database (82) prioritized the extraction of a key set
418 of variables that existed in the included trials. The harmonization process was also described as labor
419 intensive and requiring considerable clinical insight. Establishing a consistent set of categories with
420 which to code each variable, while adhering to the guiding principle of never discarding information,
421 was a major issue in the IMPACT harmonization process, as in our own. Data harmonization requires
422 solving all problems on an ongoing basis and sharing updates to the database with trialists and other
423 project collaborators.
424 Developing and validating the PROPERmed models using datasets that were originally designed and
425 conducted for many different purposes was challenging, as we were neither able to control what kind
426 of data was collected, nor how it was accomplished. However, this concern was more than offset by
427 the superior quality of the collected data, as well as the low number of missing values that is typical of
428 controlled trials.
429 In future trials on multimorbidity and polypharmacy, data collection should be standardized to a
430 greater degree, e.g., by including the agreed core outcome set proposed by Smith et al. (83). Moreover,
431 potential studies may benefit from such advance planning, as it would ensure data analyses are of high
432 quality.
433
434 **CONCLUSIONS**
435 In summary, the present manuscript presents the PROPERmed collaboration. It describes the
436 development of a database for the first IPD-MA-based prognostic model for older patients with chronic

conditions and chronic prescriptions in general practice that aims to identify patients at risk of undesirable health-outcomes. IPD-MA are becoming increasingly popular, as they have large sample sizes and employ high-quality data. However, they also require data harmonization, which is a great deal of work and may result in the loss of information.

FOOTNOTES

Funding

This work was supported by the German Innovation Fund in accordance with § 92a (2) Volume V of the Social Insurance Code (§ 92a Abs. 2, SGB V - Fünftes Buch Sozialgesetzbuch), grant number: 01VSF16018. The funding body did not play any role in the design of the study and collection, analysis, and interpretation of data, or in writing the manuscript. Andreas D. Meid is funded by the Physician-Scientist Programme of Heidelberg University, Faculty of Medicine. Rafael Perera receives funding from the NIHR Oxford Biomedical Research Council (BRC), the NIHR Oxford Medtech and In-Vitro Diagnostics Co- operative (MIC), the NIHR Applied Research Collaboration (ARC) Oxford and Thames Valley, and the Oxford Martin School. Kym Snell is funded by the National Institute for Health Research School for Primary Care (NIHR SPCR Launching Fellowship). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

Acknowledgements

The authors would like to thank all participating local data managers (Sandra Rauck, Masha Twellaar, Karin Aretz, Antonio Fenoy and Kiran Chapidi). We would also like to thank Phillip Elliott for editing the manuscript.

Author contributors

JWB, MvdA, UT, WEH, HJT, PJME, GK, JJM, DKdG, RP, PG and CM contributed to the design of the PROPERmed study. CM is the guarantor. AIG and TN wrote the first draft of the manuscript. JWB, MvdA, DBL, UT, HR, HJT, PJME, GK and CM represent the five included trials and provided all data

needed for the IPD-MA. AIG and TN developed the harmonized PROPERmed database with the support of KMAS, HR and BF. AM performed the statistical analysis with the support of RP, KS and HR. DKdG and AIG assessed the risk of bias of individual studies. All authors had full access to the data in the study and can take responsibility for the integrity of the data and the accuracy of the data analysis. The corresponding author attests that all listed authors meet the authorship criteria and that no others meeting the criteria have been omitted.

Competing interests

All authors declare no support from any additional organizations for the submitted work, no financial relationships with any organizations that might have had an interest in the submitted work in the last three years, and no other relationships or activities that could appear to have influenced the submitted work.

Ethical approval

The Ethics Commission of the Medical Faculty of the Johann Wolfgang Goethe University, Frankfurt / Main confirmed that no extra vote was necessary for the anonymous use of data in the PROPERmed IPD-MA (13/07/2017). All included trials were separately approved by their ethics commissions as follows:

ISCOPE: The Medical Ethical Committee of Leiden University Medical Center approved the study (date: 30.06.2009, reference: P09.096).

Opti-Med: The Medical Ethics Committee of the VU University Medical Centre Amsterdam approved the study (date: 12.01.2012, reference: 2011/408).

PIL: The Medical Ethics Review Board Atrium-Orbis-Zuyd approved the study (date: 15.12.2009, reference: 09-T-72 NL3037.096.09).

PRIMUM: The Ethics Commission of the Medical Faculty of the Johann Wolfgang Goethe University, Frankfurt / Main approved the study (date: 20/05/2010, reference: E 46/10).

RIME: The Ethics Commission of the University Witten / Herdecke approved the study (date: 28.02.2012, reference: 147/2011).

REFERENCES

1. Salisbury C, Johnson L, Purdy S, Valderas JM, Montgomery AA. Epidemiology and impact of multimorbidity in primary care: a retrospective cohort study. Br J Gen Pract. 2011;61(582):e12-21.
2. Fortin M, Stewart M, Poitras ME, Almirall J, Maddocks H. A systematic review of prevalence studies on multimorbidity: toward a more uniform methodology. Ann Fam Med. 2012;10(2):142-51.
3. Weng MC, Tsai CF, Sheu KL, Lee YT, Lee HC, Tzeng SL, et al. The impact of number of drugs prescribed on the risk of potentially inappropriate medication among outpatient older adults with chronic diseases. QJM. 2013;106(11):1009-15.
4. Thomsen LA, Winterstein AG, Sondergaard B, Haugbolle LS, Melander A. Systematic review of the incidence and characteristics of preventable adverse drug events in ambulatory care. Ann Pharmacother. 2007;41(9):1411-26.
5. Campbell SE, Seymour DG, Primrose WR, Lynch JE, Dunstan E, Espallargues M, et al. A multi-centre European study of factors affecting the discharge destination of older people admitted to hospital: analysis of in-hospital data from the ACMEplus project. Age Ageing. 2005;34(5):467-75.
6. Hanlon JT, Schmader KE, Ruby CM, Weinberger M. Suboptimal prescribing in older inpatients and outpatients. J Am Geriatr Soc. 2001;49(2):200-9.
7. Rodrigues MC, Oliveira C. Drug-drug interactions and adverse drug reactions in polypharmacy among older adults: an integrative review. Rev Lat Am Enfermagem. 2016;24:e2800.
8. Aarts S, Patel KV, Garcia ME, Van den Akker M, Verhey FR, Metsemakers JF, et al. Co-Presence of Multimorbidity and Disability with Frailty: An Examination of Heterogeneity in the Frail Older Population. J Frailty Aging. 2015;4(3):131-8.
9. Meid AD, Haefeli WE. Age-Dependent Impact of Medication Underuse and Strategies for Improvement. Gerontology. 2016;62(5):491-9.
10. Meid AD, Quinzler R, Groll A, Wild B, Saum KU, Schottker B, et al. Longitudinal evaluation of medication underuse in older outpatients and its association with quality of life. Eur J Clin Pharmacol. 2016;72(7):877-85.
11. Pirmohamed M, James S, Meakin S, Green C, Scott AK, Walley TJ, et al. Adverse drug reactions as cause of admission to hospital: prospective analysis of 18 820 patients. BMJ. 2004;329(7456):15-9.
12. Meid AD, Quinzler R, Freigofas J, Saum KU, Schottker B, Holleczek B, et al. Medication Underuse in Aging Outpatients with Cardiovascular Disease: Prevalence, Determinants, and Outcomes in a Prospective Cohort Study. PLoS One. 2015;10(8):e0136339.
13. Schneeweiss S, Hasford J, Gottler M, Hoffmann A, Riethling AK, Avorn J. Admissions caused by adverse drug events to internal medicine and emergency departments in hospitals: a longitudinal population-based study. Eur J Clin Pharmacol. 2002;58(4):285-91.
14. Heinrich S, Rapp K, Rissmann U, Becker C, Konig HH. Cost of falls in old age: a systematic review. Osteoporos Int. 2010;21(6):891-902.
15. Rottenkolber D, Schmiedl S, Rottenkolber M, Farker K, Salje K, Mueller S, et al. Adverse drug reactions in Germany: direct costs of internal medicine hospitalizations. Pharmacoepidemiol Drug Saf. 2011;20(6):626-34.
16. Siebenhofer A, Paulitsch MA, Pregartner G, Berghold A, Jeitler K, Muth C, et al. Cluster-randomized controlled trials evaluating complex interventions in general practices are mostly ineffective: a systematic review. J Clin Epidemiol. 2018;94:85-96.
17. Patterson SM, Cadogan CA, Kerse N, Cardwell CR, Bradley MC, Ryan C, et al. Interventions to improve the appropriate use of polypharmacy for older people. Cochrane Database Syst Rev. 2014(10):CD008165.
18. Smith SM, Wallace E, O'Dowd T, Fortin M. Interventions for improving outcomes in patients with multimorbidity in primary care and community settings. Cochrane Database Syst Rev. 2016;3:CD006560.

19. Geurts MM, Talsma J, Brouwers JR, de Gier JJ. Medication review and reconciliation with cooperation between pharmacist and general practitioner and the benefit for the patient: a systematic review. Br J Clin Pharmacol. 2012;74(1):16-33.
20. Holland R, Desborough J, Goodyer L, Hall S, Wright D, Loke YK. Does pharmacist-led medication review help to reduce hospital admissions and deaths in older people? A systematic review and meta-analysis. Br J Clin Pharmacol. 2008;65(3):303-16.
21. Johansson T, Abuzahra ME, Keller S, Mann E, Faller B, Sommerauer C, et al. Impact of strategies to reduce polypharmacy on clinically relevant endpoints: a systematic review and meta-analysis. Br J Clin Pharmacol. 2016;82(2):532-48.
22. Muth C, Beyer M, Fortin M, Rochon J, Oswald F, Valderas JM, et al. Multimorbidity's research challenges and priorities from a clinical perspective: the case of 'Mr Curran'. Eur J Gen Pract. 2014;20(2):139-47.
23. von Buedingen F, Hammer MS, Meid AD, Muller WE, Gerlach FM, Muth C. Changes in prescribed medicines in older patients with multimorbidity and polypharmacy in general practice. BMC Fam Pract. 2018;19(1):131.
24. Muth C, Uhlmann L, Haefeli WE, Rochon J, van den Akker M, Perera R, et al. Effectiveness of a complex intervention on Prioritising Multimedication in Multimorbidity (PRIMUM) in primary care: results of a pragmatic cluster randomised controlled trial. BMJ Open. 2018;8(2):e017740.
25. Marengoni A, Angleman S, Melis R, Mangialasche F, Karp A, Garmen A, et al. Aging with multimorbidity: a systematic review of the literature. Ageing Res Rev. 2011;10(4):430-9.
26. Steyerberg EW, Moons KG, van der Windt DA, Hayden JA, Perel P, Schroter S, et al. Prognosis Research Strategy (PROGRESS) 3: prognostic model research. PLoS Med. 2013;10(2):e1001381.
27. National Institute for Health Care Excellence. Multimorbidity: clinical assessment and management. National Institute for Health and Care Excellence; 2016.
28. Von Korff M, Wagner EH, Saunders K. A chronic disease score from automated pharmacy data. Journal of Clinical Epidemiology. 1992;45(2):197-203.
29. Clark DO, Von Korff M, Saunders K, Baluch WM, Simon GE. A chronic disease score with empirically derived weights. Med Care. 1995;33(8):783-95.
30. Quinzler R, Freitag MH, Wiese B, Beyer M, Brenner H, Dahlhaus A, et al. A novel superior medication-based chronic disease score predicted all-cause mortality in independent geriatric cohorts. J Clin Epidemiol. 2019;105:112-24.
31. Alonso-Moran E, Nuno-Solinis R, Onder G, Tonnara G. Multimorbidity in risk stratification tools to predict negative outcomes in adult population. Eur J Intern Med. 2015;26(3):182-9.
32. Voruganti TR, O'Brien MA, Straus SE, McLaughlin JR, Grunfeld E. Primary care physicians' perspectives on computer-based health risk assessment tools for chronic diseases: a mixed methods study. J Innov Health Inform. 2015;22(3):333-9.
33. Saver BG, Mazor KM, Hargraves JL, Hayes M. Inaccurate risk perceptions and individualized risk estimates by patients with type 2 diabetes. J Am Board Fam Med. 2014;27(4):510-9.
34. Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. BMJ. 2010;340:c221.
35. Tierney JF, Vale C, Riley R, Smith CT, Stewart L, Clarke M, et al. Individual Participant Data (IPD) Meta-analyses of Randomised Controlled Trials: Guidance on Their Use. PLoS Med. 2015;12(7):e1001855.
36. Willeboordse F, Hugtenburg JG, van Dijk L, Bosmans JE, de Vries OJ, Schellevis FG, et al. Opti-Med: the effectiveness of optimised clinical medication reviews in older people with 'geriatric giants' in general practice; study protocol of a cluster randomised controlled trial. BMC Geriatr. 2014;14:116.
37. Willeboordse F, Schellevis FG, Chau SH, Hugtenburg JG, Elders PJM. The effectiveness of optimised clinical medication reviews for geriatric patients: Opti-Med a cluster randomised controlled trial. Fam Pract. 2017;34(4):437-45.

38. Bosch-Lenders D, Maessen DW, Stoffers HE, Knottnerus JA, Winkens B, van den Akker M. Factors associated with appropriate knowledge of the indications for prescribed drugs among community-dwelling older patients with polypharmacy. Age Ageing. 2016;45(3):402-8.
39. Blom J, den Elzen W, van Houwelingen AH, Heijmans M, Stijnen T, Van den Hout W, et al. Effectiveness and cost-effectiveness of a proactive, goal-oriented, integrated care model in general practice for older people. A cluster randomised controlled trial: Integrated Systematic Care for older People--the ISCOPE study. Age Ageing. 2016;45(1):30-41.
40. O'Halloran J, Miller GC, Britt H. Defining chronic conditions for primary care with ICPC-2. Fam Pract. 2004;21(4):381-6.
41. EuroQolGroup. EuroQol - a new facility for the measurement of health-related quality of life. Health Policy. 1990;16:199-208.
42. Lutomski JE, Baars MA, Schalk BW, Boter H, Buurman BM, den Elzen WP, et al. The development of the Older Persons and Informal Caregivers Survey Minimum DataSet (TOPICS-MDS): a large-scale data sharing initiative. PLoS One. 2013;8(12):e81673.
43. von der Schulenburg J-M, Claes C, Greiner W, Uber A. The German Version of the EuroQoL Group questionnaire. Z Gesundh Wiss 1998;6(1):3-20.
44. Kaplan RM. The minimally clinically important difference in generic utility-based measures. COPD. 2005;2(1):91-7.
45. Walters SJ, Brazier JE. Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. Qual Life Res. 2005;14(6):1523-32.
46. Explained ES. International Standard Classification of Education (ISCED) 2018 [updated 28.08.2018. Available from: http://ec.europa.eu/eurostat/statistics-explained/index.php/International_Standard_Classification_of_Education_(ISCED).
47. World Health Organization. Classifications of diseases 2020 [Available from: https://www.who.int/classifications/icd/icdonlineversions/en/.
48. Lamberts H, Woods M. International Classification of Primary Care. Oxford: Oxford University Press; 1987.
49. WONCA. ICPC-2. International Classification of Primary Care. Oxford: Oxford University Press; 1998.
50. Diederichs C, Berger K, Bartels DB. The measurement of multiple chronic diseases--a systematic review on existing multimorbidity indices. J Gerontol A Biol Sci Med Sci. 2011;66(3):301-11.
51. Mukherjee B, Ou HT, Wang F, Erickson SR. A new comorbidity index: the health-related quality of life comorbidity index. J Clin Epidemiol. 2011;64(3):309-19.
52. Ou HT, Mukherjee B, Erickson SR, Piette JD, Bagozzi RP, Balkrishnan R. Comparative performance of comorbidity indices in predicting health care-related behaviors and outcomes among Medicaid enrollees with type 2 diabetes. Popul Health Manag. 2012;15(4):220-9.
53. Cheng L, Cumber S, Dumas C, Winter R, Nguyen KM, Nieman LZ. Health related quality of life in pregeriatric patients with chronic diseases at urban, public supported clinics. Health Qual Life Outcomes. 2003;1:63.
54. Charlson ME, Pompei P, Ales KL, C.R. M. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis. 1987;40(5):373-83.
55. Renom-Guiteras A, Meyer G, Thurmann PA. The EU(7)-PIM list: a list of potentially inappropriate medications for older people consented by experts from seven European countries. Eur J Clin Pharmacol. 2015;71(7):861-75.
56. O'Mahony D, O'Sullivan D, Byrne S, O'Connor MN, Ryan C, Gallagher P. STOPP/START criteria for potentially inappropriate prescribing in older people: version 2. Age Ageing. 2015;44(2):213-8.
57. Dreischulte T, Donnan P, Grant A, Hapca A, McCowan C, Guthrie B. Safer Prescribing--A Trial of Education, Informatics, and Financial Incentives. N Engl J Med. 2016;374(11):1053-64.

58. Carnahan RM, Lund BC, Perry PJ, Culp KR, Pollock BG. The relationship of an anticholinergic rating scale with serum anticholinergic activity in elderly nursing home residents. Psychopharmacol Bull. 2002;36(4):14-9.
59. Carnahan RM, Lund BC, Perry PJ, Pollock BG, Culp KR. The Anticholinergic Drug Scale as a measure of drug-related anticholinergic burden: associations with serum anticholinergic activity. J Clin Pharmacol. 2006;46(12):1481-6.
60. Hilmer SN, Mager DE, Simonsick EM, Ling SM, Windham BG, Harris TB, et al. Drug burden index score and functional decline in older people. Am J Med. 2009;122(12):1142-9 e1-2.
61. Hilmer SN, Mager DE, Simonsick EM, Cao Y, Ling SM, Windham BG, et al. A drug burden index to define the functional burden of medications in older people. Arch Intern Med. 2007;167(8):781-7.
62. Cao YJ, Mager DE, Simonsick EM, Hilmer SN, Ling SM, Windham BG, et al. Physical and cognitive performance and burden of anticholinergics, sedatives, and ACE inhibitors in older women. Clin Pharmacol Ther. 2008;83(3):422-9.
63. Duran CE, Azermai M, Vander Stichele RH. Systematic review of anticholinergic risk scales in older adults. Eur J Clin Pharmacol. 2013;69(7):1485-96.
64. Palmer M, Harley D. Models and measurement in disability: an international review. Health Policy Plan. 2012;27(5):357-64.
65. Saliba D, Elliott M, Rubenstein LZ, Solomon DH, Young RT, Kamberg CJ, et al. The Vulnerable Elders Survey: A Tool for Identifying Vulnerable Older People in the Community. Journal of the American Geriatrics Society. 2001;49(12):1691-9.
66. Isaacs B. An Introduction to Geriatrics. London: Bailliere, Tindall & Cassell; 1965.
67. Von Korff M, Ormel J, Keefe FJ, Dworkin SF. Grading the severity of chronic pain. Pain. 1992;50(2):133-49.
68. Gandek B, Ware JE, Aaronson NK, Apolone G, Bjorner JB, Brazier JE, et al. Cross-Validation of Item Selection and Scoring for the SF-12 Health Survey in Nine Countries. Journal of Clinical Epidemiology. 1998;51(11):1171-8.
69. Ware J, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. Med Care. 1996;34(3):220-33.
70. Aaronson NK, Muller M, Cohen PDA, Essink-Bot M-L, Fekkes M, Sanderman R, et al. Translation, Validation, and Norming of the Dutch Language Version of the SF-36 Health Survey in Community and Chronic Disease Populations. Journal of Clinical Epidemiology. 1998;51(11):1055-68.
71. Sheikh JI, Yesavage JA, Brooks JOr, Friedman L, Gratzinger P, Hill RD, et al. Proposed factor structure of the Geriatric Depression Scale. Spring. 1991;3(1):23-8.
72. Hoyl MT, Alessi CA, Harker JO, Josephson KR, Pietruszka FM, Koelfgen M, et al. Development and Testing of a Five-Item Version of the Geriatric Depression Scale. Journal of the American Geriatrics Society. 1999;47(7):873-8.
73. Yesavage JA, Brink TL, Rose TL, Lum O, Huang V, Adey M, et al. Development and validation of a geriatric depression screening scale: A preliminary report. Journal of Psychiatric Research. 1982;17(1):37-49.
74. Yu DS, Yan EC, Chow CK. Interpreting SF-12 mental component score: an investigation of its convergent validity with CESD-10. Qual Life Res. 2015;24(9):2209-17.
75. Vilagut G, Forero CG, Pinto-Meza A, Haro JM, de Graaf R, Bruffaerts R, et al. The mental component of the short-form 12 health survey (SF-12) as a measure of depressive disorders in the general population: results with three alternative scoring methods. Value Health. 2013;16(4):564-73.
76. Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. Annals of internal medicine. 2019;170(1):51-8.
77. Streit S, Gussekloo J, Burman RA, Collins C, Kitanovska BG, Gintere S, et al. Burden of cardiovascular disease across 29 countries and GPs' decision to treat hypertension in oldest-old. Scand J Prim Health Care. 2018;36(1):89-98.

78. Streit S, Verschoor M, Rodondi N, Bonfim D, Burman RA, Collins C, et al. Variation in GP decisions on antihypertensive treatment in oldest-old and frail individuals across 29 countries. BMC Geriatr. 2017;17(1):93.
79. Boef AG, le Cessie S, Dekkers OM, Frey P, Kearney PM, Kerse N, et al. Physician's Prescribing Preference as an Instrumental Variable: Exploring Assumptions Using Survey Data. Epidemiology. 2016;27(2):276-83.
80. den Elzen WP, Lefebre-van de Fliert AA, Virgini V, Mooijaart SP, Frey P, Kearney PM, et al. International variation in GP treatment strategies for subclinical hypothyroidism in older adults: a case-based survey. Br J Gen Pract. 2015;65(631):e121-32.
81. Schaap LA, Peeters GM, Dennison EM, Zambon S, Nikolaus T, Sanchez-Martinez M, et al. European Project on OSteoArthritis (EPOSA): methodological challenges in harmonization of existing data from five European population-based cohorts on aging. BMC Musculoskelet Disord. 2011;12:272.
82. Marmarou A, Lu J, Butcher I, McHugh GS, Mushkudiani NA, Murray GD, et al. IMPACT database of traumatic brain injury: design and description. J Neurotrauma. 2007;24(2):239-50.
83. Smith SM, Wallace E, Salisbury C, Sasseville M, Bayliss E, Fortin M. A Core Outcome Set for Multimorbidity Research (COSmm). Ann Fam Med. 2018;16(2):132-8.

Figure 1. Selection and maintenance of PROPERmed participants

HA = Hospital Admissions; IPD = Individual Participant Data; m = months; NA = Not Available

Table 1. Main characteristics of the included trials

	PRIMUM	Opti-Med	PIL	ISCOPE	RIME
Title	PRIoritizing MUltimedication in Multimorbidity	Optimized clinical Medication reviews in older people	Polypharmacy intervention in Limburg	Integrated C Care for Older People	Reduction of potentially inadequate Medication in the Elderly
Study region	Hesse, Germany	Amsterdam, The Netherlands	Southern Limburg, The Netherlands	Leiden, The Netherlands	Witten and Hanover, Germany
Start/End	2010-2012	2013-2015	2010-2014	2009-2011	2012-2014
Design	2-arm parallel cRCT	2-arm parallel cRCT	2-arm parallel cRCT, stepped wedge	2-arm parallel cRCT	2-arm parallel cRCT
Setting	General practices (n=72)	General practices (n=22)	General practices (n=24)	General practices (n=51)	General practices (n=139)
Study population	N=502 ≥60 years ≥5 chronic prescriptions ≥3 chronic conditions	N=518 ≥65 years ≥1 chronic prescription Geriatric giant	N=727 ≥60 years ≥5 chronic prescription	N=1,617 ≥75 years complex problems on ≥3 domains	N=1,197 ≥6 chronic prescriptions
Exclusion criteria	Life expectancy of ≤12 months Abuse of alcohol or illegal drugs Inability to fill in questionnaires and to participate in telephone interviews Dementia	Life expectancy of ≤6 months Severe psychiatric illness or other reasons why GP regards patient as unable to participate Recent medication review (last 6 months) Dementia	Life expectancy of ≤12 months Abuse of alcohol or illegal drugs Dementia	Life expectancy of ≤3 months or terminal illness Dementia	Life expectancy of ≤6 months as assessed by the treating GP Dementia
Intervention	Structured medication review including several intervention components (complex intervention)	Structured medication review including several intervention components (complex intervention)	Structured medication review including several intervention components (complex intervention)	Integrated care plan using a functional geriatric approach	Structured medication review including several intervention components (complex intervention)
Data collection (study visits)	0, 6, 9 months	0, 3, 6 months	0, 3, 6, 12 months	0, 6, 12 months	0, 6, 12 months
Data collection (method)	CRF, patient questionnaire, telephone interview	CRF, patient questionnaire	CRF, patient questionnaire, home visit	Patient questionnaire, home visit	CRF, patient questionnaire, telephone interview

cRCT = cluster-randomized controlled trial; CRF = Case Report Form; GP = General Practitioner.

Table 2. Data harmonization

Type of variable	Variable	Harmonization	PRIMUM	Opti-Med	PIL	ISCOPE	RIME
Sociodemographic and lifestyle-related	Age	No need of harmonization	-	-	-	-	-
	Sex	Recoding of labels	-	-	-	-	+
	Living situation	Recoding of categories and labels	+	+	+	+	+
	Educational level	Recoding of categories and labels	-	-	+	+	+
	Smoking status	Recoding of categories and labels	+	x	+	x	+
Morbidity-related	Single conditions	Clinical harmonization	+	+	+	+	+
	Disease count[a]	Creation of new variable	+	+	+	+	+
	Multimorbidity index[b]	No need of harmonization	-	x	x	x	-
Medication-related	Medication inappropriateness and omissions[c]	Creation of new variable	+	+	+	+	+
Functional status- and well-being-related	Depressive symptoms	Dichotomization based on cut-offs	+	+	+	+	+
	Functional status	Standardized mean differences	+	+	+	+	+
	Quality of life (EQ-5D-3L index)	No need of harmonization	-	-	-	-	-
	Health-related quality of life comorbidity index	Creation of new variable	+	+	+	+	+
	Hospital admissions	Recoding of categories and labels	+	+	x	x	+

Legend: + data was harmonized; **-** no need to harmonized; **x** systematic missing

[a] modified Diederichs list; [b] modified CCI; [c] modified EU-PIM list, STOPP-START criteria, Dreischulte high risk prescribing list, ADS, DBI, Anticholinergic Drug Burden

Table 3. Proxies for multimorbidity instruments

	Diederichs (50)	PCS (51-53)	MCS (51-53)	Charlson (54)
Cancer	X			X
Chronic obstructive pulmonary disease / asthma	X	X	X	X
Coronary heart disease	X[a]	X	X	X
Depression	X		X	
Diabetes	X	X	X	X
Degenerative neurological disorders (Parkinson)		X	X	
Hearing problems	X			
Heart failure	X		X	X
HIV infection /AIDS			X	X
Hypertension	X	X	X	
Osteoarthritis	X[b]	X	X	
Osteoporosis	X			
Rheumatoid/seropositive arthritis		X	X	
Cerebrovascular disease	X	X	X	X
Vision problems	X	X		
Total number of diagnoses included in the original list	17	20	15	17
Total number of diagnoses included in PROPERmed (% from the total number in the original list)	12 (71 %)	12 (60 %)	6 (40 %)	7 (41 %)

[a] Includes heart disease, myocardial infarction and angina pectoris from the original list.
[b] Includes rheumatoid / seropositive arthritis.

Table 4. Risk of bias assessment using the PROBAST tool*

Domain / Overall judgement	RoB / Applicability	(a) Assessment for each trial					(b) Combined assessment of all trials
		PRIMUM	**Opti-Med**	**PIL**	**ISCOPE**	**RIME**	
Domain 1: Participant selection	RoB	Low	Unclear [a]	Low	Low	Low	Low [e]
	Applicability	Low concern	Low concern	Low concern	Low concern	Low concern	Low concern
Domain 2: Predictors	RoB	Low	Low	Low	Low	Low	Unclear [b]
	Applicability	Low concern	Low concern	Low concern	Low concern	Low concern	Low concern
Domain 3: Outcomes	RoB	Low	Low	Low	Low	Low	Low
	Applicability	Low concern	Low concern	Low concern	Low concern	Low concern	Low concern
Domain 4: Sample size and participant flow	RoB	Not applicable to single studies, only overall judgement possible					Not available [c]
Domain 5: Analysis	RoB	Not applicable to single studies, only overall judgement possible					Not available [c]
Overall judgment [d]	**RoB**	**Low**	**Unclear [a]**	**Low**	**Low**	**Low**	**Unclear**
	Applicability	**Low concern**	**Low concern**	**Low concern**	**Low concern**	**Low concern**	**Low concern**

RoB = Risk of Bias

[a] Unclear because analyzed participants had a different state of health at baseline.

[b] Unclear due to uncertainty about the definition and assessment method of potential predictor variables across included studies.

[c] Data not available at the time the RoB assessment was performed.

* Adapted from PROBAST tool" (Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. Ann Intern Med [Internet]. 2019;170(1):51)

Table 5. Descriptives of outcomes and candidate prognostic variables

Outcomes (six months follow-up)

	Variables	Categories	PRIMUM	Opti-Med	PIL	ISCOPE	RIME	PROPERmed
Outcomes	Quality of life: EQ-5D-3L, Index value – Mean (SD)		0.78 (0.24)	0.74 (0.21)	0.76 (0.22) [a]	0.69 (0.26) [a]	0.73 (0.25)	0.73 (0.25)
	Hospitalization (at least one) – Number (%)		79 (16.5%)	66 (16.3%)	NA[b]	NA[b]	259 (23.9%)	404 (20.5%)

Candidate prognostic variables – Sociodemographic and lifestyle-related (baseline)

	Variables	Categories	PRIMUM	Opti-Med	PIL	ISCOPE	RIME	PROPERmed
Sociodemographic and lifestyle-related	Age – Mean (SD)		72.18 (6.86)	77.24 (7.86)	73.18 (7.53)	81.83 (5.00)	76.90 (4.93)	77.56 (7.04)
	Sex (female) - Number (%)		262 (52.2%)	334 (65.0%)	356 (47.4%)	1,110 (68.9%)	599 (50.0%)	2,661 (58.2%)
	Living situation - Number (%)	Living at home	496 (98.8%)	486 (98.6%)	698 (95.0%)	1,471 (92.2%)	1,154 (96.4%)	4,305 (94.9%)
		Institutionalized living	1 (0.2%)	27 (1.4%)	37 (5.0%)	125 (7.8%)	43 (3.6%)	233 (5.1)
	Educational level - Number (%)	Low	311 (62.8%)	114 (23.6%)	526 (71.2%)	861 (53.3%)	219 (18.3%)	2,031 (44.9%)
		Medium	145 (29.3%)	225 (45.7%)	126 (17.1%)	616 (28.7%)	882 (73.7%)	1,994 (44.0%)
		High	39 (7.9%)	152 (30.7%)	87 (11.8%)	129 (8.0%)	96 (8.0%)	503 (11.1%)
	Smoking status - Number (%)	Smoker	248 (50.9%)	NA	146 (29.4%)	NA	543 (45.4%)	937 (43.0%)
		Ex-smoker	193 (39.6%)		284 (51.1%)		583 (48.7%)	1,060 (48.6%)
		Non-smoker	46 (9.4%)		67 (13.5%)		71 (5.9%)	184 (8.4%)

Candidate prognostic variables – Morbidity-related (baseline)

	Variables	Categories	PRIMUM	Opti-Med	PIL	ISCOPE	RIME	PROPERmed
Morbidity-related	Single conditions - Number (%)	Hypertension	422 (84.1%)	295 (57.4%)	441 (62,9%)	793 (49.7%)	1,054 (88.1%)	3,005 (66.6%)
		Diabetes	267 (53.2%)	115 (22.4%)	291 (41.5%)	398 (24.9%)	547 (45.7%)	1,618 (35.9%)
		Coronary heart disease	215 (42.8%)	127 (24.7%)	293 (41.8%)	388 (24.3%)	515 (43.0%)	1,538 (34.1%)
		Osteoarthritis	268 (53.4%)	160 (31.1%)	186 (26.5%)	366 (22.9%)	542 (45.3%)	1,522 (33.7%)
		Vision problems	72 (14.3%)	190 (37.0%)	215 (30.7%)	407 (25.5%)	573 (47.9%)	1,457 (32.3%)

		Chronic obstructive pulmonary disease	146 (29.1%)	109 (21.2%)	147 (21.0%)	285 (17.8%)	276 (23.1%)	963 (21.3%)
		Hearing problems	13 (2.6%)	47 (9.1%)	105 (15.0%)	249 (15.6%)	471 (39.3%)	885 (19.6%)
		Cancer	47 (9.4%)	137 (26.5%)	91 (13.0%)	311 (19.5%)	236 (19.7%)	822 (18.2%)
		Heart failure	86 (17.1%)	53 (10.3%)	61 (8.7%)	207 (13.0%)	403 (33.7%)	810 (18.0%)
		Cerebrovascular disease	98 (19.5%)	79 (15.4%)	111 (15.8%)	278 (17.4%)	165 (13.8%)	731 (16.2%)
		Osteoporosis	53 (10.6%)	98 (19.1%)	108 (15.4%)	167 (10.5%)	259 (21.6%)	685 (15.2%)
		Depression	85 (16.9%)	46 (8.9%)	53 (7.6%)	138 (8.6%)	143 (11.9%)	465 (10.3%)
		Rheumatoid arthritis	25 (5.0%)	15 (2.9%)	40 (5.7%)	64 (4.0%)	224 (18.7%)	368 (8.2%)
		Parkinsonism	7 (1.4%)	22 (4.3%)	6 (0.9%)	41 (2.6%)	26 (2.2%)	102 (2.3%)
		HIV/AIDS	0	1 (0.2%)	0	0	0	1 (0.02%)
	Disease count according to modified Diederichs – Median (IQR)		4 (1)	3 (2)	3 (2)	2 (2)	4 (2)	3 (2)
	Modified Charlson comorbidity index – Median (IQR)		3 (2)	NA	NA	NA	3 (3)	3 (3)

Candidate prognostic variables – Medication-related (baseline)

	Variables	**Categories**	**PRIMUM**	**Opti-Med**	**PIL**	**ISCOPE**	**RIME**	**PROPERmed**
Medication-related	No. of drugs† - Median (IQR)		9 (3)	5 (4)	7 (3)	5 (5)	9 (4)	7 (5)
	No. of Potentially Inappropriate Medications (PIM) according to the modified EU-PIM list - Median (IQR)		1 (1)	1 (1)	1 (2)	1 (1)	1 (2)	1 (1)
	STOPP criteria (modified) – Median (IQR)		2 (1)	1 (1)	1 (1)	1 (1)	2 (1)	1 (1)
	START criteria (modified) – Median (IQR)		1 (2)	1 (2)	1 (2)	1 (2)	1 (2)	1 (2)
	Dreischulte criteria (modified) – Median (IQR)		0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
	Drug Burden Index (DBI, modified) - Median (IQR)		0 (1)	0 (1)	0 (1)	0 (1)	0 (1)	0 (1)
	Anticholinergic Drug Burden (ADB) according to Duran - Median (IQR)		0 (1)	0 (1)	0 (1)	0 (1)	0 (1)	0 (1)
	Anticholinergic Drug Scale (ADS) according to Carnahan - Median (IQR)		0 (1)	0 (1)	1 (2)	1 (1)	1 (1)	1 (1)

Candidate prognostic variables – Functional status and well-being-related, plus hospital admissions (baseline)

	Variables	Categories	PRIMUM	Opti-Med	PIL	ISCOPE	RIME	PROPERmed
Functional status and well-being-related	Functional status and frailty – Mean (SD)		2.83 (2.81) [c]	5.04 (2.83) [d]	1.71 (2.44) [e]	3.13 (3.01) [e]	2.86 (2.57) [c]	
	Depressive symptoms (binary) – Number (%)	GDS 15/5	73 (14.8%) [k]	NA	NA	233 (15.2%) [k]	183 (15.3%) [k]	489 (15.2%)
		SF 12/36	NA	148 (30.9%) [j]	33 (4.9%) [j]	NA	298 (28.0%) [j]	479 (21.6%)
	Depression (Score) – Mean (SD)	GDS 15/5	2.39 (2.29) [k]	NA	NA	2.16 (2.50) [k]	0.73 (0.94) [k]	1.67 (2.17)
		SF 12/36	NA	63.87 (22.83) [j]	76.85 (17.85) [j]	NA	54.03 (9.50) [j]	63.21 (18.73)
	Pain (binary) – Number (%)		439 (87.5%) [f]	372 (75.0%) [g]	537 (77.8%) [h]	801 (50.3%) [i]	888 (74.9%) [h]	3,037 (68.0%)
	Pain Intensity – Mean (SD)		1.23 (0.65) [f]	1.15 (0.79) [g]	1.18 (0.77) [h]	NA	NA	1.19 (0.74)
	Pain Disability – Mean (SD)		0.60 (0.79) [f]	0.94 (0.61) [j]	0.78 (0.65) [j]	NA	1.07 (0.75) [j]	0.89 (0.73)
	Pain SF – Mean (SD)		NA	2.39 (1.05) [j]	2.18 (1.10) [j]	NA	2.74 (1.39) [j]	2.5 (1.27)
	Modified health-related quality of life comorbidity index, mental – Median (IQR)		2 (2)	1 (2)	1 (2)	1 (2)	2 (2)	1 (3)
	Modified health-related quality of life comorbidity index, physical – Median (IQR)		8 (4)	4 (4)	5 (4)	4 (4)	7 (5)	5 (5)
	Quality of life: EQ-5D-3L, Index value (baseline) – Mean (SD)		0.79 (0.23)	0.74 (0.21)	0.77 (0.21) [a]	0.68 (0.27) [a]	0.73 (0.26)	0.73 (0.25)
Hospital admissions	Hospitalization (at least once in past 12 months) – Number (%)		81 (16.1%)	NA	574 (78.1%)	412 (25.8%)	435 (36.4%)	1,502 (37.3%)

Legend: †No. of drugs refer to ATC codes (i.e., single active ingredients or fixed combinations as listed in ATC version 2012).
NA = not available.
[a] EQ-5D was used in an experimental version in PIL and ISCOPE; [b] at 12 months from baseline; [c] VES-13; [d] Scale for Mobility problems; [e] Katz-15; [f] von Korff (cut-off ≥ 1 for categorical variable); [g] VAS (cut-off ≥ 1 for categorical variable); [h] PIL SF-36 and RIME SF-12; [i] single question; [j] Opti-Med and RIME MCS score SF-12 (cut-off < 50.2 for categorical variable), and PIL MCS score SF-36 (cut-off < 42 for categorical variable); [k] PRIMUM and ISCOPE GDS 15 (cut-off ≥ 5 for categorical variable), RIME GDS 5 (cut-off ≥ 2 for categorical variable).

Table 6. Knowledge gaps addressed by the PROPERmed collaboration

A harmonized individual participant data database of older patients with multimorbidity and polypharmacy in general practice
A prognostic model for deterioration in health-related quality of life in older patients with multimorbidity and polypharmacy in general practice
A prognostic model for hospital admissions in older patients with multimorbidity and polypharmacy in general practice