15 Usability Recommendations for Delivering Clinical Guidelines on Mobile Devices

James Mitchell
Keele University
j.a.mitchell@keele.ac.uk

Ed de Quincey
Keele University
e.de.quincey@keele.ac.uk

Charles Pantin
UHNM NHS Trust
c.pantin@keele.ac.uk

Naveed Mustfa
UHNM NHS Trust
naveed.mustfa@uhnm.nhs.uk

Local point of care clinical guidelines exist in numerous formats and cover a variety of clinical information, normally created on a national and local level. They are generally available as basic web pages, PDFs or documents. Despite widespread availability and use, accessing clinical guidelines and information can be highly inefficient and restrictive. This reflective study investigates the evaluation of a clinical guidelines mobile application in the challenging area of co-design with clinicians. It aimed to answer if the selected methods of user centred design were suitable when working with limited access to users and what design recommendations can be elicited/changed by utilising user centred design (UCD) methods to gather feedback on features and functions. Specifically, this study utilised a mixed-method UCD approach and triangulation technique (Think-aloud and idea writing, screen recording and system usability scale). This culminated into the creation of 15 recommendations for developing clinical guidelines applications for mobile devices.

User centred design. Clinical guidelines. Mobile application design.

1. INTRODUCTION

Clinical guidelines are produced by numerous organisations, health trusts and hospitals worldwide (NICE, 2020; Pantin et al., 2006). They exist in numerous formats and cover a variety of clinical information (NICE, 2020; Pantin et al., 2006). Point of care clinical guidelines (patient treatment and medical process information designed for use at the point of care) are generally available as basic web pages, PDFs or documents (NICE, 2020; Pantin et al., 2006). Clinicians require agile access to these guidelines and an efficient delivery method (Free et al., 2013; Takeshita et al., 2002). Despite widespread availability and use however, accessing clinical guidelines and information can be highly inefficient and restrictive (Burton and Edwards, 2019; Littlejohns et al., 2003).

A previous study by the authors (Mitchell et al., 2020) investigated this issue by producing and evaluating a clinical guidelines app (based on the Bedside Clinical Guidelines (BCG)) utilising user-centred design methods (Abras et al., 2004; Norman, 1986; usability.gov, 2019). The main aim of the research was to identify and evaluate suitable methods for presenting clinical guidelines on a mobile phone interface, with a focus on efficiency and usability. The results from this study were then used to create a set of recommendations for developing mobile device apps to deliver clinical guidelines.

A total of thirteen (n=13) recommendations were developed:

- Cross-Platform
- List view with A to Z and Categories
- Basic filter
- Easy access menu (such as tabbed)
- Minimise manual tasks (e.g. Manual calculations)
- Minimise the requirement to use other systems (if possible), e.g. if a drug dosage calculation is required, this should be available to the clinician without the need to use another app or system. This may not be possible due to security, organisational governance or limitations of technology.
- Decision algorithms to be displayed in-line with the guideline information
- The original 'flowchart' decision algorithm is provided
- Minimise the number of warnings/alerts to avoid 'alert fatigue'
- Acronym use is prevalent in medicine, but not all clinicians have knowledge of acronyms. Methods to address both experts and novices should be adopted.
- Warnings should be more explicit and adopt better salience for the user
- Guideline sentences should be reduced

- Content Pages should utilise icons/images as well as headers

A second iteration of the application was developed implementing the 13 recommendations listed (see figure 1). This paper presents the evaluation and adaption of these recommendations through UCD, as well as reflecting on the UCD process and related methods, and their appropriateness for user groups with limited availability.

A mixed-methods UCD approach has been used based on the triangulation technique (Heale and Forbes, 2013; Noble and Heale, 2019) represented in *figure 2*. This enabled qualitative and quantitative data collection to inform design recommendations. The methods (Think-aloud and idea writing, screen recording and the system usability scale), rationale for selection and results are discussed in the following sections in more detail (2 - 5)
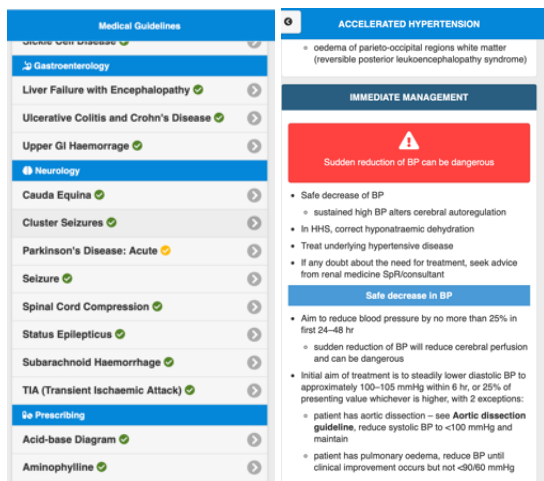

Figure 1: *Example presentation of clinical information*

1.1 Ethics

Ethical approval was granted by Keele University Research Governance in the Faculty of Natural Sciences (ERP2370) and from Research and Development at the University Hospitals of North Midlands NHS Trust. A letter of access was provided for the duration of the study.

2. THINK-ALOUD

The think-aloud (Nielsen, 1992) technique was chosen to elicit feedback as it provided a method of understanding how users navigated the structure of the BCG app as well as their thoughts during the process of using it to complete basic clinical information retrieval scenarios. This method also allowed for the discovery of usability issues during information retrieval which may not have been identified during other methods of testing (i.e., focus groups).
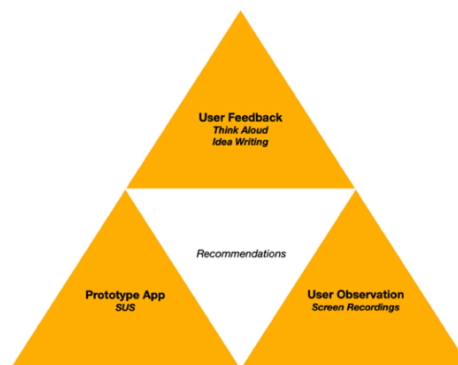

Figure 2: *Triangulation techniques used to evaluate the application*

2.1 Recruitment

Participants for this study were recruited via invitation emails which were sent via the Year Four Medical Lead to all fourth-year medical students at Keele University School of Medicine. In total, 38 students were recruited.

Participants
Participants were selected using a convenience sampling method and were required to have some medical knowledge. As access to clinicians was severely limited, it was decided that fourth year medical students would provide the adequate medical knowledge required for the basic information retrieval tasks and be accessible to the researcher. Demographics were not collected as this data would not provide relevant information in terms of design and implementation feedback. The purpose was to test the usability of the BCG app and therefore, selection based on demographics would not offer any further information required in terms of usability feedback.

2.2 Protocol

One to one sessions of fifteen minutes were arranged with all respondents who were offered certificates of participation. Participants were greeted and a brief overview of how think-aloud sessions are conducted was provided to allow users to understand the purpose and process of the session. Participants did not have access to the BCG app prior to the session and they were also not provided information on how the BCG app functions.

Participants were then asked to follow a process containing basic clinical scenarios. used to emulate clinical workflow. Research by Tu et al. (Tu et al., 2004) discusses the modelling of clinical guidelines for integration into clinical workflow and shows how a clinical workflow can be modelled using clinical scenarios. This is echoed in other studies such as Cossu et al. (Cossu et al., 2014), and UK based studies by Payne et al. (Payne et al., 2014) and Kwa (Kwa et al., 2015, 2014) where clinical scenarios were utilised during initial testing. The

processes shown in these studies were used to inform the design of the clinical scenarios used for this study, described further in this section.

Screen and audio recordings were made of each think-aloud session using Apple QuickTime 10.4 and an iPhone X running iOS 13 tethered via USB.

2.3 Session overview

Clinical Scenarios
For this study, the clinical scenarios were developed with assistance of a Lead Respiratory Consultant at the Royal Stoke University Hospital. They were developed to ensure participants accessed specific guidelines and utilised guideline components such as text, warnings and decision algorithm tools.

For all sessions, the following process was followed:

- Participants were provided with an overview of how to open the BCG app (the app was not opened at this stage).

- Participants were provided with the three basic clinical scenarios. The clinical scenarios were based on three information retrieval tasks:

    (a) In the subsequent management of Unstable Angina, what is the recommended dose and method of administering Aspirin?
    (b) During fluid management in Acute Heart Failure, when should an echocardiogram be sought?
    (c) In the management flowchart of Hyperkalaemia, what is the recommended action where Plasma K+ 6.0-6.4 mmol/L and Acute ECG changes are present?

Scenario (a) was design to ask participants to retrieve basic text-based information. Scenario (b) was created to ask participants to retrieve text information contained in a warning, this enabled the analysis of how clinicians interact with the warnings contained in the BCG app. Scenario (c) was created to ask participants to retrieve information contained within a decision algorithm, again allowing analysis of how participants use the inline algorithm tools.

1. Audio and device screen recording was started (this is utilised for analysis discussed in section 6)

2. Participants were asked to access the BCG app and retrieve the required information (via scenarios) whilst discussing their actions and thoughts. Participants were asked to clarify comments during the session.

3. After completing the basic clinical scenarios, participants were given a brief

demonstration of other features in the app e.g. Acronym support and Calculation tools.

4. Participants were asked to complete an SUS questionnaire (discussed in section 4).

During the think aloud, prompting questions were utilised when specific feedback was required but did not naturally occur during the session i.e. where a participant describes something as good they would be asked to elaborate and explain why it is "good". These questions related to aspects such as design, layout, content and usability.

2.4 Think-aloud Analysis

Data Analysis
To identify themes from the think aloud session, audio recordings were transcribed verbatim and analysed by the primary researcher. User actions during screen recordings were analysed and coded.

Overall coded theme and category analysis

Table 1: *Main themes identified during the Thematic Analysis*

Theme	Description
Main Menu	*App content page*
Guideline Layout	*Design of the guidelines including Typeface/Colour, how the information is presented*
Warnings/Alerts	*Presentation and content of warnings/alerts contained within the guidelines*
Decision algorithm	*Presentation and content of decision algorithms contained within the guidelines*
Filter function	*Presentation and content of filter functions contained within the guidelines and on the main menu*
Features or functions not present	*Suggestion or requirement of features and functions that are not currently available in the BCG app*

Table 2: *Number of comments related to each theme*

Main Menu	*9*
Guideline Layout	*94*
Warnings/Alerts (not specific to task)	*52*
Flowchart/Decision Algorithm tool	*74*
Text/font/colour	*7*
Filter Function	*16*

A total of 252 comments were coded over the 38 sessions analysed. In some cases, comments were considered neutral or irrelevant and therefore excluded from the final analysis. Examples include comments where participants would discuss unrelated information such as medical knowledge not relevant to the scenario or BCG app. On average participants made seven (n=7) comments that were coded/themed with a range of three to

eighteen (3 – 18). Six themes were identified during the analysis shown in table 1. For each participant, an average of 3 themes were identified with a range of two to five (2 – 5).

Of the six identified in table 1, the themes most discussed (both positive and negative) were *GUIDELINE LAYOUT* (37.3% of comments) and *DECISION ALGORITHMS* (29.4% of comments). Details of the number of comments for each theme are provided in table 2. From overall comments, *GUIDELINE LAYOUT* and *DECISION ALGORITHM* represented a combined total of 66.6% (n=168/252). WARNINGS also represented a large portion of comments (20.6% of comments).

Comments for each theme (Table 1) were also categorised in terms of the categories which described comments overall. Similar studies have categorised in this way (Li et al., 2012) as it enables an understanding of the proportion of comments for each theme in terms of how they relate to aspects such as usability, clinical workflow etc. Therefore simplifying the identification of comments related to factors such as the content of the guideline versus the design on the guideline. Of the six themes identified (Table 1) four categories were created to code each comment (Table 3). Table 3 identifies these categories and provides a description of each. Of the four categories identified in Table 3, the categories most discussed were *USABILITY* (56% of comments) and *VISIBILTY* (23% of comments). From overall comments, usability and visibility represented a combined total of 79% of all comments (n=199/252).

Table 3: *Categories of coding and description of each category*

Category	Description
Usability	Comments which are considered to refer to how the app is used, how the information can be accessed and how the users 'feel' in terms of its use. (e.g. "I like how this looks")
Visibility	Comments which refer to the visibility, colour, salience, layout etc. (e.g. "I didn't notice it because it didn't stand out")
Clinical Workflow	Comments specifically refer to use of the app and its functions in wards/hospitals (e.g. "this would be really useful when treating patients as it can get busy on the wards")
BCG Content	Comments which specifically refer to the content itself – including text, knowledge and specific medical information/methods (e.g. "I would have expected this section to be above investigations")

Participant comment analysis

Sessions were also analysed for consistent patterns in how participants utilised features of the BCG app. The following sections discuss the results of each particular theme presented in table 2 and provides examples of comments made by participants in relation to each.

2.5 Main Menu

All participants navigated the main menu without the need for prompting or further instruction. All participants were able to access the specific guidelines. In some cases, they utilised the filter function (n=30) – this is analysed further in this section. Some participants made specific positive comments in relation to the use of icons and headers for the sections provided. This can be summarised by the following participant quote:

> "that's nice that you have this at the beginning so that you could flick through and see just an overview of all the things that you have on it"

Of the nine (n=9) comments made by participants in reference to the main menu, seven (n=7) were considered positive and two (n=2) were considered negative. An example of a positive comment referenced the use of categories:

> "You've got headings which I like"

The majority of positive comments reference the layout and ease of use in terms of finding what they need. An example of a negative comment mentioned the following:

> "maybe it'd be nicer if it was just the big blue header and then you can open and close"

The negative comments (n=2) in reference to the main menu all have similar themes in terms of presenting the content in an accordion type (open and close) view, as the above comment suggests.

2.6 Guideline Layout

The majority of respondents made general comments regarding the layout of the guidelines. A total of 94 comments were coded in reference to the guideline layout, a large proportion of all comments that were coded (37.3% of all comments). Of that total, 69 were considered positive and 25 were considered negative. As the following example highlights, most positive comments referenced the ease of finding information or the clarity of the layout:

> "I think just how it's laid out signs and symptoms and then investigations and then differential diagnosis. I feel like it's laid out in a good order and there's not too much text as well. Cause I find

that when I'm using NICE and stuff like that, there's so much text."

In terms of negative comments, the majority of participants suggested a more collapsible layout may be beneficial. One user did specifically mention that in one of the guidelines, scrolling was undesirable. The participant stated:

"I think it's a bit long to like scroll down on set. I think just separating it a bit and bit might be a bit useful".

Other comments suggested that there should be an overview of all the content (e.g. a content section or titles at the top of each guideline) to facilitate user understanding of the guideline layout:

"Maybe like at the top there could be like a mini, like contents where you could click on, for example, subsequent management and anything"

Feedback also suggested that the order of the content would be more beneficial if different from its current layout, for example:

"my only sort of thought with that is having the differentials above investigations. So as you read an investigations, you already know what really not helped."

2.7 Warnings/Alerts

A large proportion of respondents specifically mentioned the layout of warnings or gave specific feedback regarding the information contained in the BCG app warnings (n=52/252, 20.63% of all comments). Of all the comments coded to particular themes, warnings/alerts received the majority of negative comments (45% of all negative comments). This was due to participants expecting the use of acronyms or shortened versions such as 'ECG' or 'ECHO'. This was evident through comments such as:

"So you would expect acronyms to be in there too"

"It was more because I didn't see that it was anything to do with an echo"

Some participants suggested the information should be repeated in context within the guidelines. Summarised by some participant in the following comments:

"So I was expecting it to be in the standard text. Um, I normally would have looked at that… Perhaps a repeat of that. So, repeating the warning, in the information."

"I was actually looking for a bullet that said echocardiogram. Okay. Um, so perhaps you

could include it as both. It's like in the red and as a bullet point."

Some participants also suggested warnings that contained too much text were harder to assimilate when scrolling through the BCG guidelines. In reference to the amount of text contained in a warning, one participant mentioned:

"I like things that are bullet pointed and then inset bullet point, and then the detailing."

In some cases, negative comments were associated with users not finding the information contained within the warning. Whilst some suggested that the information should be repeated, other users specifically mentioned that they felt the medical procedure would not necessarily be presented in a warning box, as the comment below suggests:

"I think I just assumed. That, that wouldn't be. I didn't read that. I don't know why, although it looks like it's designed to be more important. I guess I assumed that an echo wouldn't be that important"

However, other participants suggested that the information in refence to an echocardiogram would not necessarily be expected to be in a section with fluid management.

"so maybe it's just me missing it. And then if we hadn't, since it's about an echocardiogram, …put that in the fluid management"

These comments echo other participant comments referencing the repeat of information in the main text. It also highlights individual user behaviour and how participants assimilate the information contained in the guideline. One participant specifically mentioned their workflow may have contributed to them missing the information contained within the warning:

"I'm so used to just looking straight at the text rather than in boxes. Um, and usually I go back to boxes to see if things are important. Yeah. Um, but I'm usually, yeah, that's hard to get straight to text, so that's why I missed it"

The majority of positive comments referred to the salience of the warning, in particular the use of colour. Participant specifically mentioned the warning salience during the sessions:

"I definitely saw like the red warning thing, so I guess that is quite, it shows that it's important. I guess if it's immediate, that means that you probably want to put at the top, which you guys did and. This pops out because you don't see this kind of thing on the other, on the other one that I saw"

"That's quite nice to have like a big warning to make sure that you do what you need to do"

"Cause it's an, a red box with a warning and like, I think anyone would automatically look and make sure like, what's that warning about"

2.8 Decision Algorithm

All sessions were analysed in terms of how the participants interacted with the decision algorithm tool. The users had two options in terms of how to access the flowchart information they required to complete the scenario, a programmatic version and an image of the original flowchart/algorithm. However, they were not made aware of this in order to assess which method they would instinctively access. t is worth note that the decision algorithm tool is more salient in terms of design than the button to access the original version (figure 3). However, participants had access to both programmatic and original version of the decision algorithm within the same area of the guideline (Figure 3). Table 4 shows the number of participants utilising each version.
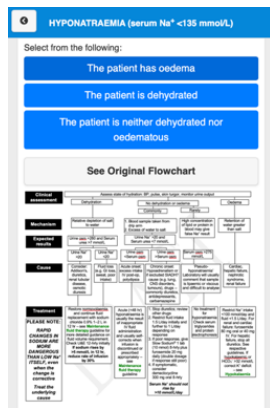

***Figure 3**: Programmatic and original flowchart (inline)*

***Table 4**: Number of participants utilising each version of the decision algorithm*

Utilised programmatic version	Utilised original version
37	1

Of all participants (n=38), all but one (n=37) accessed the programmatic version of the decision algorithm. Comments made by participants on the design and use of the inline decision algorithms were overwhelmingly positive. Of the seventy-four (n=74) comments made by participants in reference to the tool, sixty-seven (n=67) were positive and seven (n=7) were classified as negative. Specifically, one participant mentioned when comparing the two decision algorithms:

"so this is just a different way of presenting that digital flow chart. I think I liked the other (ref to new method) because this is too complicated (ref to original). And I think when needed quickly on the ward and you want to see something that probably not the best way".

Another participant also reflected on the design, specifically stating:

"it helps you follow in your head. I find that flowcharts can be a bit much sometimes following it. Whereas this specifically just gives you the answer you need rather than everything on stuff. So, it makes it a bit easier to follow and easy to get the information you need".

One participant also discussed the decision algorithm. Directly referencing the amount of information presented and reflecting on the need for specific information. This was also reflected in their comment, were they stated:

"sometimes when it's like branching and you having to look everywhere to find exactly what you need, it's to the point"

Interestingly, there appeared to be a separate viewpoint on the use of information for learning as opposed to clinical use. This was highlighted specifically by one participant in reference to the presentation of the original decision algorithm (flowchart), stating

"I guess the original flow chart be good for learning".

2.9 Filter Function

Participant screen recordings were analysed to determine if any utilised the filter function (Figure 4), both on the main menu and within the guideline (Table 5).
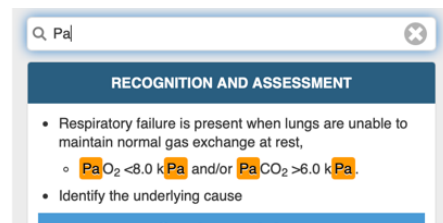

***Figure 4**: Filter function available in each guideline and via the main menu*

***Table 5**: Number of times participants accessed the filter function*

Utilised filter in the Main Menu	Utilised filter in the guideline
18	12

As the table highlights, participants accessed the filter function during the session with no prompting or instruction. The main menu filter was accessed by eighteen of the participants (n=18/38), and the guideline filter function was accessed by twelve participants (n=12/38). Participants also specifically mentioned using the filter function during use, describing it as a "quicker" or "faster" method of retrieving information. Overall, 16 comments were

coded in reference to the filter functionality of the BCG app. Of these comments, 14 were considered positive, with general positive comments as mentioned. In terms of negative comments, 2 were identified by separate users. One user specifically mentions in terms of clinical workflow the following:

"If you didn't know, you could type in potentially the symptoms or to go into cardiac"

Another participant also suggested that the filter function may be more useful if it allows the user to:

"move to the next part"

This suggests that the user is navigated to each highlight of the filter in a similar method that some PDF/Browser word filters function.

Features or functions not present
Participants mentioned aspects of clinical information that may be useful within the bedside clinical guidelines. In particular, drug calculation tools or information on specific treatments. As the scope of this study is to investigate the delivery of existing guidelines, it is beyond the scope of this study to investigate information on guidelines that do not currently exist. However, it is interesting to highlight that the information needs of participants does differ especially in terms of clinical expertise and interest.

Positive/Negative analysis
As well as identifying themes and categories, sessions were also analysed in terms of whether comments were positive or negative. This allows for an overall analysis of participants attitude towards the BCG app and enables the identification of specific features/themes that participants described in negative or positive terms. The following describes how comments were coded as positive or negative:

- A *positive* reaction or general comment (e.g. "this is really great" or describing the use of a feature in a positive manner (e.g. "This would be really useful when…")
- A *negative* reaction or general comment (e.g. "I don't like this..") or any criticism, suggestion of alternative methods or ways in which the user prefers (e.g. "this is good but I would like it if it did…"

Each coded comment considered negative or positive was analysed by theme. Overall, of the 252 comments coded, a total of 182 were coded positive and 70 were coded negative. The majority of coded comments considered positive (n=182/252 or ~82%) focussed on *GUIDELINE LAYOUT* and the *DECISION ALGORITHMS*, both of which, as mentioned, received the most comments overall. Interestingly, the majority of negative coded

comments also focussed on *GUIDELINE LAYOUT* (36% of all negative comments). However, this was most likely due to the high number of comments received overall. *WARNINGS/ALERTS* (46% of all negative comments) received a greater proportion of negative comments relative to overall comments. Of the 52 comments referencing warnings/alerts, 32 were coded negative and 20 coded positive.

Errors/Issues
The think-aloud sessions were also analysed for any occasions were participants encountered issues or errors. Table 6 describes the three areas created to describe the issues found.

***Table 6:*** *Issue types and descriptions*

Issue/error Type	Description
Information retrieval issue	*Unable to retrieve the necessary information to complete the scenario or where the user selects the wrong information.*
Usability issue	*Interacts with the app in a way they perceive negative due to its design or functionality*
Other	*Discovers a bug or app issue not related to information retrieval or usability*

Table 7: *Number of occurrences of each issue during think-aloud sessions*

Issue/error Type	Number of occurrences
Information retrieval	*18*
Usability issue	*9*
Other	*1*

A total of 26 issues/errors occurred over the 38 sessions, 68% of sessions. The 26 issues occurred in 21 sessions of the 38 with a range of 0 – 2 issues per session. Table 7 provides an overview of the types of issues and the number of occurrences for each type. Of the 18 occurrences of issues related to information retrieval, 9 occurrences were related to participants locating information incorrectly. Despite the scenario specifically asking users the 'dose of aspirin in subsequent management', 9 participants provided the initial dose contained in the management section. When prompted to locate the information in 'subsequent management' some users did state that an overview of the sections available in the guideline may be useful. This was highlighted in the comments contained in the Guideline layout section. The 9 other occurrences of information retrial errors all related to user not able to locate information contained in the warning box provided in the Acute Heart Failure guideline. This was due to the expectation of acronyms/short

versions and the expectation of text contained in the warning would be repeated or available in the main guideline text, as mentioned in Warning/Alerts section. Of the 9 occurrences of usability issues, 2 occurrences related to locating the decision algorithm tool. During the first 3 sessions, the decision algorithm had to be activated by clicking the start button. After the initial usability issues this was changed to be inline without requiring activation, see figure 5. No further occurrences of this issue occurred in the remaining 36 sessions. This highlighted a clear usability issue that was resolved.
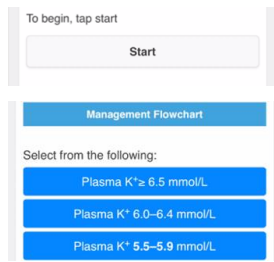
A(Top): Decision algorithm tool that requires activation
B(Bottom): Inline decision algorithm tool requires no activation

Figure 5: *Display changes for decision algorithm tools*

The most prevalent usability issue was related to users mistaking a header for a button (Figure 5 shows the header and button). 5 participants (13.5% of participants) attempted to click the header for the tool before realising the tool was already present in the guideline. This represents 56% of the usability issues identified. Upon analysing the screen recording, all 5 participants failed to scroll down far enough to visibly see the tool, therefore assumed they could activate it using the header. This could also have been caused by the gap between the header and tool (see Example B), which does not conform with best practice (Wagemans et al., 2012). Most users acknowledged the error and, on some occasions, mentioned that this would not occur after they have become more familiar with how the BCG app works. Other usability issues included an occasion where one participant could not initially locate the 'Acute heart failure' guideline in the main menu, caused by the participant looking for heart failure and did not expect 'acute' to precede the title. Another issue identified was related to the filter function within the guideline. One participant attempted to move to the next guideline by searching for it in the filter tool, this was corrected by the participant without any interjection. A further issue was identified during the 18th think-aloud session. A bug was identified where the warnings did not display when using the filter function, this was categorised as an 'other' issue as it was not specifically related to usability or information retrieval. This was fixed before further sessions were conducted. Analysis of previous 17 sessions did not identify any other occurrences of this issue and the

issue did not contribute to any negative comments or other issues identified during the previous sessions.

3. SUS

The System Usability Scale (SUS) (Brookes, 1996) was used to establish the usability level of the application from the clinicians' viewpoint. It also provided a baseline to measure future changes in the design and how they impact the usability. All participants (n=38) were asked to complete the SUS questionnaire post think-aloud session. Results were then analysed and compared to results of a previous study which had investigated the usability of a previous version of the app (Mitchell et al., 2020). This allowed for analysis of changes made to the application based on recommendations derived from previous UCD studies. These are discussed in previously published work (Mitchell et al., 2020).
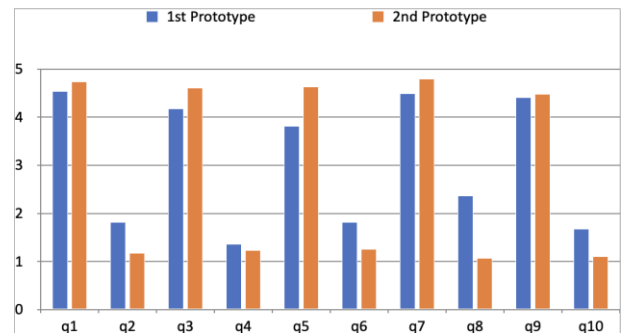
3.1 SUS results

Figure 6: *A comparison of SUS results for the 1st and 2nd application iteration*

The app was shown to maintain a high usability score, with an overall score of 93.6 out of 100 (calculated utilising the methods described in (Brookes, 1996)). This result was higher than SUS scores discussed in previously published work (Mitchell et al., 2020), 81 out of 100. The consistent results in all sessions highlight a general consensus amongst participants that they highly rate the usability of the BCG app. This was also reflected in the positive comments/feedback discussed in the think aloud sessions. A comparison of the first and second application iteration is provided in figure 6.

4. IDEA WRITING

To further evaluate the app, an 'idea writing' session (Austin, 1994; VanGundy, 1984) was conducted. As discussed in a previous publication (Mitchell et al., 2020), the concept of using the idea writing methodology was due to the necessity to elicit information in a limited time. As access to clinicians is limited, idea writing allowed for a focus group to feedback based on a 'closed' method.

4.1 Idea Writing method

During this session, clinicians interacted with the application and were asked to feedback on each aspect of the design, which was presented as a 'concept'. Although this limited open discussion (by design), it allowed for more specific feedback regarding the design of the BCG app.

Participants
This session was conducted at the Wythenshawe Hospital, part of the Manchester University NHS foundation Trust. The session was conducted with four (n=4) participants. Participants were selected using the convenience sampling method.

4.2 Feedback

Feedback provided during the idea writing session was largely positive. Specifically, participants used words such as "very useful" and "good" to positively describe the app, an example includes:

> "simplify the content as too wordy to be used in emergency although info all good - my suggestion is to use flowcharts as much as possible as first thing you see then have the fuller content below or linked to separate page"

Although this is related the authoring of the guidelines, this does specifically mention the need for succinct information delivery in an emergency, specifically delivery utilising the decision algorithm. Another comment refers to guideline titles:

> "simplify and lose acute from the section titles as it makes it harder to search for subjects"

Although this has only mentioned by a single participant during the focus groups and think aloud sessions, an interesting point was raised regarding succinct information and how it is displayed in the content pages. It also has similarities to a usability issue which occurred during the think aloud sessions where a user was unable to locate the Acute Heart Failure guidelines because it was superseded by the word acute. Another comment also referenced the layout of guidelines, specific to warnings:

> "warnings at top of pages"

This was in contrast to feedback received during other focus groups and think aloud sessions. However, it does highlight that individual preference may be a key factor in delivering clinical information and this requires further investigation.

5. OVERVIEW OF RESULTS

Table 8 provides an overview of the main findings presented in sections 2-5. Each finding is presented with the method utilised and how it has affected the recommendations presented in the introduction of this paper.

Table 8: Overview of method findings and outcomes

#	Finding	Method	Outcome
1	Inline decision tools caused less errors	Think aloud and video analysis	**Adapting** existing recommendations to include inline activation
2	Warning design should be more explicit and salient	Think aloud and Idea Writing	**Adapting** existing recommendations to include explicit, salient warnings
3	Warning text be repeated to avoid missing critical information	Think aloud	Adding a new recommendations based on repeat warning text.
4	Easier to find guidelines if unnecessary wording is removed	Think aloud and idea writing	Adding a new recommendations based on removing wording from titles in content pages

6. DISCUSSION

This study utilised a mixed-method triangulation approach to inform the improvement of a mobile application for delivering bedside clinical guidelines. The use of the think-aloud technique with clinical scenarios and the 'idea writing' focus group, as well as the SUS methodology produced data which has informed on the impact of implementing recommendations and identified clear usability issues (i.e. decision algorithm activation). Despite the overlap in the findings of these methods, unique insights were elicited from participants. These methods also enabled evaluation of a clinical application where access to relevant users (clinicians) is extremely limited and restricted in terms of time. They also offer a unique insight into the use of these techniques as no studies that have combined these techniques to inform the delivery of bedside clinical guidelines could be found. The evaluation has provided a number of specific and general findings relevant to the development of the BCG app. In terms of layout, some participants referred to the order of content and specified alternative ordering. This is indicative of how preferences differ between individuals. Similar findings were also discussed in a previous publication, where personal preference has contributed to a large amount of variation in the apps clinicians utilise. This is further impacted by the requirements of the delivered information in terms of educational use as opposed to clinical use. Participants conveyed the need for a more in-depth delivery of information when learning. This is

highlighted in Karen Davies's review on the information-seeking behaviour of doctors, which states two main behaviours when clinicians are seeking information, one seeking facts and another seeking literature (Davies, 2007). This also reflects the findings of the observational study (Mitchell et al., 2020), which found that Junior clinicians appear to use technology to establish knowledge which requires more information. Senior clinicians utilise technology for knowledge affirmation. The use of acronyms also suggests there are differences in the needs of individual clinicians from a knowledge perspective. Interestingly, the topic of warnings generated much discussion in terms of the information they contain. Specifically, the use of acronyms was expected by participants which is in direct contrast to feedback received during previous sessions (Mitchell et al., 2020). This may be due to the subject matter utilised within the warning. The scenarios utilised echocardiograms, a subject the participants were familiar with. It remains to be seen if other more complex subjects and less used acronyms would highlight knowledge gaps. However, previous findings highlighted the need to provide both acronyms and explications.

7. CONCLUSIONS

This study aimed to answer if the selected methods of user centred design were suitable when working with limited access to clinicians. Based on the feedback received and the adaption of recommendations, these methods have worked efficiently on providing feedback and evaluation for an app. The study also aimed to evaluate what design recommendations can be elicited/changed by utilising user centred design methodologies. The evaluation of the thirteen recommendations during this paper suggests that at least two of the original recommendations discussed in the introduction of this paper need to be adapted (as presented in Table 8): Decision algorithms and Calculation tools should be displayed in-line with the guideline information, clearly outlined to distinguish from the main content, and ready to be used (i.e., does not require activation); Warnings should be succinct, explicit and adopt a salient design to ensure visibility. The findings also suggest the addition of two new recommendations , they are as follows: Text contained in alerts or warnings should also be available within the text it refers to; Remove unnecessary wording in titles e.g. Instead of 'Acute Heart Failure' use 'Heart Failure'. The adaption of previous recommendations and the addition of new recommendations has culminated in the creation of 15 recommendations for developing clinical information delivery applications for mobile devices. The following provides an overview of the final set of recommendations.

1. Be cross platform
2. Provide multiple methods of accessing content in list views (i.e., A to Z and Categories)
3. Minimise unnecessary wording in titles (i.e., 'Acute heart failure' should be presented as 'heart failure'
4. Have a menu that can be easily accessed, preferably using a tabbed menu design
5. Utilise icons/images as well as headers
6. Provide a basic filter function to filter content in both menu and information sections
7. Minimise manual tasks (i.e., Drug dose calculations)
8. Provide as many tools and resources as possible to minimise the requirement to use other systems
9. Provide clear decision algorithms and calculation tools in line with content, and ready to use (i.e., does not require activation)
10. Provide original content for any tools or decision algorithms (i.e., An original flow chart)
11. Utilise acronyms, but also provide a method of understanding acronyms where possible
12. Minimise the number of warnings/alerts to avoid 'alert fatigue'
13. Display warnings/alerts in line with content, ensuring they are salient in design and succinct and explicit in content
14. Repeat warning content within the main information
15. Reduce the use of long sentences and provide information as succinctly as possible

Aside from the recommendations elicited from feedback and evaluation, it is also clear that further investigation into personalised delivery is required. Although a limited number of participants specifically mentioned layout, the feedback during the evaluation of both BCG apps highlights the eclectic nature of information delivery that satisfies user preference.

7.1 Limitations

It is worth note that the higher SUS score could be attributed to the fact that clinical students were utilised, a group that are familiar with mobile devices and clinical application use (Mitchell et al., 2020; Payne et al., 2012; Prescott et al., 2017). However, students participate in clinical practice through their university course, a requirement for all student clinicians in their final years of study. It is suspected that although this may have some effect on the results, it would not have a considerable impact as student and junior clinicians were utilised in the earlier SUS sessions and focus groups.

8. REFERENCES

Abras, C., Maloney-Krichmar, D., Preece, J., 2004. User-Centered Design.

Austin, M., 1994. Needs Assessment By Focus Groups. American Society for Training and Development.

Brookes, J., 1996. SUS: a "quick and dirty' usability Scale, Usability Evaluation In Industry. https://doi.org/10.1201/9781498710411-35

Burton, Z., Edwards, H., 2019. A little less conversation, a little more high impact action. Futur Healthc J 6, 201–201. https://doi.org/10.7861/futurehosp.6-1-s201

Cossu, F., Marrella, A., Mecella, M., Russo, A., Kimani, S., Bertazzoni, G., Colabianchi, A., Corona, A., Luise, A.D., Grasso, F., Suppa, M., 2014. Supporting Doctors through Mobile Multimodal Interaction and Process-Aware Execution of Clinical Guidelines, in: 2014 IEEE 7th International Conference on Service-Oriented Computing and Applications (SOCA). p. 183 190. https://doi.org/10.1109/soca.2014.44

Davies, K., 2007. The information-seeking behaviour of doctors: a review of the evidence. Health Information & Libraries Journal 24, 78–94. https://doi.org/10.1111/j.1471-1842.2007.00713.x

Free, C., Phillips, G., Watson, L., Galli, L., Felix, L., Edwards, P., Patel, V., Haines, A., 2013. The Effectiveness of Mobile-Health Technologies to Improve Health Care Service Delivery Processes: A Systematic Review and Meta-Analysis. Plos Med 10, e1001363. https://doi.org/10.1371/journal.pmed.1001363

Heale, R., Forbes, D., 2013. Understanding triangulation in research. Évid Based Nurs 16, 98. https://doi.org/10.1136/eb-2013-101494

Kwa, A., Carter, M., Page, D., Wilson, T., Brown, M., Baxendale, B., 2015. NOTTINGHAM UNIVERSITY HOSPITAL GUIDELINES APP– IMPROVING ACCESSIBLITY TO 650 HOSPITAL CLINICAL GUIDELINES. Proceedings of the International Conference on Ergonomics & Human Factors 2015 220.

Kwa, A., Wilson, T., Carter, M., Page, D., Brown, M., Bennett, O., Miles, G., Baxendale, B., 2014. 0162 Developing A Mobile App To Improve Access And Application Of Key Hospital Guidelines. Bmj Simul Technology Enhanc Learn 1, A25.2-A25. https://doi.org/10.1136/bmjstel-2014-000002.60

Li, A.C., Kannry, J.L., Kushniruk, A., Chrimes, D., McGinn, T.G., Edonyabo, D., Mann, D.M., 2012. Integrating usability testing and think-aloud protocol analysis with "near-live" clinical simulations in evaluating clinical decision support. Int J Med Inform 81, 761–772. https://doi.org/10.1016/j.ijmedinf.2012.02.009

Littlejohns, P., Wyatt, J.C., Garvican, L., 2003. Evaluating computerised health information systems: hard lessons still to be learnt. Bmj 326, 860 863. https://doi.org/10.1136/bmj.326.7394.860

Mitchell, J., Quincey, E. de, Pantin, C., Mustfa, N., 2020. The Development of a Point of Care Clinical Guidelines Mobile Application Following a User-Centred Design Approach. Lecture Notes in Computer Science 294–313. https://doi.org/10.1007/978-3-030-49757-6_21

NICE, 2020. NICE [WWW Document]. NICE Clinical Guidelines. URL : https://www.nice.org.uk/guidance (accessed 9.2.20).

Nielsen, 1992. Evaluating the thinking aloud technique for use by computer scientists. Advances in Human-Computer Interaction Vol. 3 69–82.

Noble, H., Heale, R., 2019. Triangulation in research, with examples. Évid Based Nurs 22, 67. https://doi.org/10.1136/ebnurs-2019-103145

Norman, D.A., 1986. User Centered System Design. https://doi.org/10.1201/b15703

Pantin, C., Mucklow, J., Rogers, D., Cross, M., Wall, J., Partnership, T.B.C.G., 2006. Bedside clinical guidelines: the missing link. Clin Med 6, 98–104. https://doi.org/10.7861/clinmedicine.6-1-98

Payne, K.F., Weeks, L., Dunning, P., 2014. A mixed methods pilot study to investigate the impact of a hospital-specific iPhone application (iTreat) within a British junior doctor cohort. Health Inform J 20, 59 73. https://doi.org/10.1177/1460458213478812

Payne, K.F.B., Wharrad, H., Watts, K., 2012. Smartphone and medical related App use among medical students and junior doctors in the United Kingdom (UK): a regional survey. Bmc Med Inform Decis 12, 121. https://doi.org/10.1186/1472-6947-12-121

Prescott, O., Millar, E., Nimmo, G., Wales, A., Edgar, S., 2017. 21st century medical education: critical decision-making guidance through smartphone/tablet applications—the Lothian pilot. Bmj Simul Technology Enhanc Learn 3, 60–64. https://doi.org/10.1136/bmjstel-2016-000157

Takeshita, H., Davis, D., Straus, S.E., 2002. Clinical Evidence at the Point of Care in Acute Medicine: A Handheld Usability Case Study. Proc Hum Factors Ergonomics Soc Annu Meet 46, 1409–1413. https://doi.org/10.1177/154193120204601601

usability.gov, 2019. User Centred Design, US Government [WWW Document]. URL https://www.usability.gov/what-and-why/user-centered-design.html

VanGundy, A.B., 1984. BRAIN WRITING FOR NEW PRODUCT IDEAS: AN ALTERNATIVE TO BRAINSTORMING. J Consum Mark 1, 67–74. https://doi.org/10.1108/eb008097

Wagemans, J., Elder, J.H., Kubovy, M., Palmer, S.E., Peterson, M.A., Singh, M., Heydt, R. von der, 2012. A Century of Gestalt Psychology in Visual Perception: I. Perceptual Grouping and Figure–Ground Organization. Psychol Bull 138, 1172–1217. https://doi.org/10.1037/a0029333