http://dx.doi.org/10.14236/ewic/HCI2022.28

Applying recommendations and working with clinical experts to understand, adapt, verify, and validate clinical information for mobile delivery

J. Mitchell
School of Computing and Mathematics
Keele University
j.a.mitchell@keele.ac.uk

E. de Quincey
School of Computing and Mathematics
Keele University
e.de.quincey@keele.ac.uk

C. Pantin
University Hospital North Midlands NHS Trust
c.pantin@keele.ac.uk

N. Mustfa
University Hospital North Midlands NHS Trust
naveed.mustfa@uhnm.nhs.uk

This paper discusses working with single clinical experts to apply a set of recommendations to adapt and verify clinical information, specifically, clinical guidelines for mobile (smartphone) applications. The paper discusses user-centred design methods to understand and evaluate the impact of guideline changes based on recommendations. It also discusses the impact of working with a clinical expert, and the validation/verification of information. Finally, it compares the outputs of these processes using a set of app recommendations from the Royal College of Physicians. The paper highlights the impact of applying app recommendations and evidences the importance and impact of working with single experts.

Single Expert, Clinical Information, User-centred Design, Usability, Clinical Guidelines

1. INTRODUCTION

Clinical knowledge is immensely complex (Gorry et al. 1978) and has a high rate of change (Wyatt and Spiegelhalter 1991). This is further complicated by factors such as synonyms and abbreviations which can be ambiguous. It is therefore important to involve clinicians in the design of systems and applications. However, limited access to clinicians and the need to have input from clinicians and clinical knowledge make it necessary to look at single person or expert studies (Yin 2009). Razak, Hanis and Dix discuss the value of single person studies where researchers can build up a relationship with a single user, leading to a deeper understanding of information or user needs that may otherwise appear irrelevant (Abdul Razak and Dix 2010). It also allows access to expertise that would otherwise require far more planning and availability. Research discussed in previous publications (Mitchell et al. 2020, 2021) has benefited from the guidance and experience of clinicians to produce a Bedside Clinical Guidelines (BCGs) mobile application. The BCGs are an example of clinical guidelines that are designed for use at the point of care. The BCGs have supported care at the bedside since 1996 and are currently utilised across 14 NHS Trusts throughout the UK, and aim to provide “consistent, evidence-based management of patients in acute hospital settings” (Pantin et al. 2006) for ’in the moment’ bedside use. The Bedside Clinical Guidelines (BCGs) deliver explicit clinical knowledge to support the development of tacit knowledge, however the understanding of how this knowledge is applied and developed is a key aspect. The support of a clinical guidelines expert ensures that the guidelines can be re-authored to consider knowledge and clinical practice. In particular, the re-authoring of complex medical guidelines make it necessary to involve a clinical expert. Typically, user-centred design (UCD) methodology recommends multiple user input (Nielsen and Landauer 1993; Maguire 2001) . The authors of this paper produced previous research (Mitchell et al. 2021) which discuss the evaluation of recommendations in terms of user interaction and usability. However, it is necessary to understand and evaluate the impact of these recommendations in terms of clinical information delivery. This paper

© *Mitchell et al. Published by*
BCS Learning and Development Ltd.

discusses the utilisation of a single clinical expert to adapt, verify, and validate clinical information for mobile (smartphone) delivery. It also discusses the impact of applying a set of recommendations based on previous research (Mitchell et al. 2021).

2. METHODOLOGY

This study utilised several methods of applying recommendations and working with a clinical expert. This section discusses the methodologies utilised in applying recommendations and the methodologies utilised with the single clinical expert.

2.1. Re-authoring

The clinical guidelines expert was provided with the recommendations elicited during the UCD processes discussed in previous research publications (Mitchell et al. 2021). Each guideline contained in BCG Medical Guidelines were then subjected to a re-authoring process by the clinical guidelines expert and then converted for use in the BCG app. To measure the impact the recommendations have had in authoring the guidelines, forty-six (n=46) clinical guidelines from the BCGs medical guidelines were analysed pre and post re-authoring. The guidelines were originally written in Word (.docx) format by numerous clinical guideline authors and then re-authored with the clinical guideline expert, taking into account the recommendations identified in a previous publication. The guidelines selected (n=46) were those that had been re-authored at the time of writing. It is worth note that the remaining guidelines contained within the BCG medical guidelines are structured and written in similar ways. The selection however provides a broad range of guidelines. The aim of this study was to:

- Analyse the original word count
- Analyse the reduction of tables and flowcharts
- Analyse the introduction of the new decision algorithms in each guideline

These aims are useful to consider for a variety of reasons. Tables can pose usability issues on mobile (smartphone) devices and presenting flowcharts or decision algorithms can be problematic in terms of mobile design constraints. The reduction of word count, or adapting the guidelines to be more succinct, has been evidenced to improve overall efficiency in terms of access to clinical information (Brumley et al. 2006).

2.2. Warning classification

Discussions were conducted with a single clinical expert on how warnings could be classified, the results of this are presented in the results section of this paper. As mentioned in previous publications, it was evident that warnings needed to be reduced to avoid alert fatigue. It was also evident that warnings needed be classified in terms of hierarchy (i.e. importance and nature). During the discussion it was determined that clinical information contained withing the warnings could be classified in terms of their relevant subject. To determine how they could be classified. A repeated single criterion card sort methodology (Rugg and McGeorge (1997)) was utilised as it provided a method of determining both the categories and how each warning could be categorised and sorted. The participant was provided with an overview of how card sorting is conducted and an example was provided using LEGO^{TM} bricks. The participant was then asked to sort each brick into various categories (e.g. size, shape, colour). This gives the participant a foundation of understanding for how card sorting is conducted, as evidenced by Rugg and McGeorge as they suggest a 'toy example' for instructing participants (Rugg and McGeorge 1997). An example of the LEGO^{TM} bricks used is provided in Figure 1.

Figure 1: *Example of LEGO^{TM} bricks used for instructing participants regarding Card Sorting*

2.3. Information verification

As the guidelines are adapted to conform to the recommendations set out in previous papers, it is imperative the information is validated (Ventola 2014). This will ensure patient safety and validate the correctness of the guideline information, necessary for CE certification and clinical user trust (Shekelle et al. 2001; Lewis and Wyatt 2014). To this end, a

verification process was established to ensure the information being converted from the re-authored .docx files to the BCG app format was correct. A simple text comparison tool was utilised to highlight inconsistencies in the two versions of text. As the process of converting the .docx files to a format compatible with the BCG app is manual, a simple verification stage was added to ensure correctness. The tool was developed using HTML, CSS and JavaScript. The tool predominantly utilises jsdiff, a JavaScript tool based on the O(ND) Difference Algorithm and its variations (Myers, 1986). The main method of assessing the accuracy of the BCGs converted for use in the BCG app was using manual validation with a clinical expert. A proof-reading method is utilised to verify each guideline. The method requires the researcher to read original documents aloud as the clinical guideline expert checks the software system version - this is completed in person or via Zoom/FaceTime. Automated methods could be investigated for future work but are out of scope for this project. Figure 2 shows an overview of the process used.

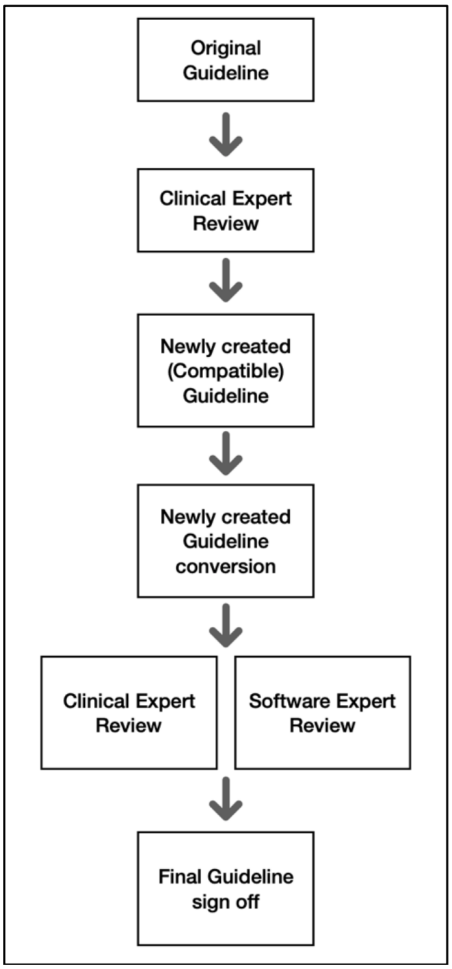

Figure 2: *Overview of the manual adaption and verification process involving a clinical expert*

2.4. Comparing the BCG app (inc converted guidelines) with current recommendations

The need for validation and information verification is an important aspect of clinical software. Wyatt et al. state the vast majority of medical apps remain without any form of regulation or safety check, and some of these may present a patient safety or other risk (Wyatt et al. 2015). In this 2015 paper Wyatt et al. introduced the Royal College of Physicians (RCP) Health Informatics Unit checklist. The Wyatt study developed and piloted an 18-item checklist to help clinicians assess the structure, functions, and impact of medical apps. The checklist assess the app internals in terms of development, the apps functionality, and if the app alleviates a problem. The checklist is outlined in Figure 3. This checklist was utilised to ensure changes to the app, guidelines, and work with the clinical expert did not impact on the app conforming with already existing recommendations. A comparison of these recommendations and the app is provided in the results section of this paper.

App name and version: _____ For iPhone / Android / other:_____

Date of filling out this checklist: _____

1. **Who developed the app, and what's inside it?**
   a) Is it clear who this app is for and how it should be used? Yes / No / Don't know
   b) Is it clear which problem the app is designed to alleviate or what outcome it helps to promote? Yes / No / Don't know
   c) Do the app developer and sponsor seem well informed about this problem or outcome, and likely to be unbiased in their approach to it? Yes / No / Don't know
   d) Have they located sound, relevant, up-to-date evidence, images, video etc to use in their app? Yes / No / Don't know
   e) Do the app screens look well designed, is text clear? Not applicable / Yes / No / Don't know
   f) Is it clear what data the app needs from the user with units defined, out of range detection and a 'clear last patient' button? Not applicable / Yes / No
   g) Does the app collect any identifiable patient information? Yes / No / Unclear
   h) Does it seem to keep user and patient data secure and private? Yes / No / Don't know
   i) If the app is designed to support any medical task,* is it CE marked? Not applicable / Yes / No / Unclear
2. **How well does the app work?**
   a) Is the app fast and easy to use in clinical settings? Yes / No / Don't know
   b) Does the app give the user usable answers or advice quickly? Not applicable / Yes / No / Unclear
   c) Do the answers, advice or calculated risks appear to be correct? Yes / No / Unclear
   d) Is there a way to feed back user comments to the app developer? Yes / No / Don't know
3. **Is there any evidence that the app does actually alleviate the problem?**
   a) Have any studies been carried out to measure the impact of using the app on clinical or patient knowledge, actions or (preferably) patient outcomes? Yes / No / Don't know
   b) Were these studies independently conducted, well designed, large enough, and applicable to the user? Not applicable / Yes / No / Don't know
   c) Did any study also examine health resource use, potential harms caused by the app, or quantify cost effectiveness? Not applicable / Yes / No / Don't know
   d) Overall, do the benefits of using this app seem likely to outweigh inconvenience and costs to the user? Yes / No / Don't know
   e) Is there any specific clinical scenario or patient subgroup in which using the app seems particularly likely to be useful? Yes - Which? / No / Unclear

Figure 3: *App Checklist developed by Wyatt et al, 2015*

3. RESULTS

3.1. Word count reduction

The results of the word count reduction are available in Table 1. Of the forty-six (n=46) guidelines analysed, nine (n=9) had an increased word count, the contributing factors to this are examined in the discussion section of this paper. Overall, the mean word count was reduced from 1045.52 to 797.7. This was an overall average reduction of 247.82, or 23%. In some guidelines (e.g. Control of hyperglycaemia in the ill patient) the word count was reduced by over

1000 words. It is important to note that the words or information reduced in each guideline were not deemed vital by the clinical expert. Each guideline's author is being consulted as part of the process to ensure robustness and quality. The word reduction is part of a mobile first approach in the design.

3.2. BCG Component analysis

There is also evidence of a substantial reduction in the number of tables (Table 2). The number of tables in the guidelines analysed was originally 42. After re-authoring, the number of tables was reduced by 36 to 6 tables, a reduction of 83%. Analysis shows that the original versions of the guidelines contained fifteen flowcharts or manual decision algorithms, this was reduced to 0. In terms of the new decision algorithms (DAs), 32 were created in the guidelines analysed. These are comprised of either original decision algorithms being converted to the new format or text information that is not deemed to be required for immediate access (i.e. requires the clinician to complete steps). As shown in Table 3, the original set of guidelines analyses contained warnings 130. After re-authoring, the number of warnings was reduced to 28. This results in a reduction of 78%. As already mentioned, these warnings were not deemed vital by the clinical expert. These changes have been implemented based on following the recommendations set out in chapter 6 (section 6.8). The reduction in word count can be linked with the recommendations of reducing long sentences, providing information as succinctly as possible, and minimising the number of warnings/alerts to avoid 'alert fatigue'. Although not a recommendation in terms of the 15 outlined in previous research (Mitchell et al. 2021), the reduction of tables is also a factor in the reduction of text. It is worth note that the word count reduction is despite the introduction of decision algorithms and calculation tools.

3.3. BCG Warning classification Results

380 warnings were identified in an analysis of the BCG Medical Guidelines. The single expert was asked to sort a selection of 20 warnings selected at random from the BCG. Each of the warnings used are provided in Table 4. Warnings were tagged with an identifying number (i.e. which warning it was from the 380 identified), the warning text, which section of the medical guidelines it is taken from and the page number. These allowed each to be identified in terms of classification after the card-sorting session. The results of the card sorting produced several categories. The clinical expert suggested to sort into Risk, workflow, and finally subject. The initial card sort produced two categories (n=2). The second card sort expanded these to three categories (n=3). The fourth card sort produced four new categories (n=4). Finally, the fourth card sort produced six categories (n=6), although not all cards could be sorted into these categories. Each of the card sorting rounds were recorded. The results are provided in Tables 6 to 9.

3.4. Application of the card sorting categories to all warnings

This aspect of the research was designed with the aim to pilot if card sorting would be a useful tool in providing categories and hierarchy to the 380 warnings contained within the BCG Medical Guidelines. The single repeated criterion card sorting method produced criterion such as 'risk level', 'management' or 'diagnosis'. In further sessions conducted as part of a single person study with a clinical expert, these criteria were applied to clinical warnings contained within the BCG guidelines. During these sessions the clinical expert categorised 100 warnings over three sessions. The warnings were categorised based on the card sorting categories produced in sort 2, shown in Table 5, as this seemed to be the most relevant to implement. During this session, it was suggested by the clinical expert that some warnings should also be contained in a further 'VERY HIGH RISK' category and as such the coding session was adapted to include this. Each warning was also coded based on the categories derived during the 4th card sort, shown in Table 9. Table 10 shows the results of the initial analysis of the 100 coded warnings. The majority (n=43) were coded as Intermediate level, with 25 and 20 warnings coded as high and low respectively. Few warnings were classified as VERY HIGH (n=5). Of the 100 coded warnings (1-100 of 380 warnings in the BCG Medical Guidelines), 7 were deemed inappropriate and therefore not categorised. On attempting to categorise warnings in terms of the categories presented in Table 31, it was apparent that the time required would be unsuitable for the limited availability of the clinical expert. At this stage, the clinical expert categorised the warnings over a longer period of time (approx. 3 weeks). During this category analysis, it was evident further categories of analysis were required. The clinical expert also classified each of the warnings (if possible) in terms of their likelihood of occurrence, in terms of illness or diagnosis, and the severity. The clinical expert analysed method and process:

Risk

Problem: If we classify risk, will people ignore lower level? A risk of '100 chance of occurring' will still means 1 in 100 patients will be harmed

A risk ranking of a warning box might be classified in two parts:

Table 1: *Word count analysis after editing of guidelines*

Guideline	Original Word Count	Word count after re-authoring	% Reduction
Standard infection prevention measures	814	620	24
Hand hygiene	741	525	29
Use of personal protective equipment	850	528	38
Screening for MRSA/SA and MGNB/ESBL/CPE	1105	184	83
Management of hospitalised patients with MRSA	493	493	0.00
Management of patients with ESBL/MGNB	13	13	0.00
Clostridium difficile infection (CDI)	1230	534	57
HIV infection testing	1204	1059	12
Sepsis management	898	515	43
Acute hot joint, septic arthritis and gout	1182	908	23
Cellulitis	904	688	24
Community acquired meningitis	381	559	-47
Fever in the returning traveller	989	919	7
Neutropenic sepsis	1105	989	11
Triage of patients with hyperglycaemia	181	181	0.00
Control of hyperglycaemia in the ill patient	1738	1034	41
Diabetic ketoacidosis and hyperosmolar hyperglycaemic state	1277	1420	-11
Fluid deficit/maintenance management flowchart	1257	487	61
Maintenance fluid therapy	864	725	16
Fluid resuscitation	1209	629	48
Unstable angina	1077	1059	2
Acute myocardial infarction	2002	2037	-2
Thoracic aortic dissection	922	1111	-21
Cardiac tamponade	324	333	-3
Acute heart failure	3020	1758	42
Cardiac arrhythmias	1442	1453	-1
Atrial fibrillation	1134	1300	-15
Infective endocarditis	1800	1022	43
Spontaneous pneumothorax	511	347	32
Acute severe asthma in adults	1106	1035	6
Exacerbation of chronic obstructive pulmonary disease (COPD)	872	765	12
Community-acquired pneumonia	1811	1022	44
Hospital-acquired pneumonia	1343	796	41
Respiratory failure	855	134	84
Pleural infection and empyema	630	527	16
Pleural effusion – investigation of	406	602	-48
Accelerated (malignant) hypertension	918	814	11
Delirium (acute confusional state) in older people	2250	1234	45
Hypothermia in older people	753	672	11
Management of constipation in hospitalised elderly patients	325	532	-64
Management of falls in A&E and wards	1007	734	27
Transient loss of consciousness (blackout/syncope)	883	781	12
Bleeding disorders in adults	1361	974	28
Management of sickle cell disease	2182	1731	21
Management bleeding in patient on dabigatran or rivaroxaban	229	501	-119
Spontaneous leucopenia or thrombocytopenia	496	410	17

Table 2: *Component numbers in the BCGs.*

Tables	42
Tables Removed	36
Remaining tables	6
New DAs	32
Flowcharts	15

Table 3: *Warning numbers in the BCGs.*

# Warnings	130
# Warnings after re-authoring	28

1. If clinician does not follow advice in the box, likelihood of event happening
2. Severity of event

These could be classified by analysis looking at the evidence. At the very least, speciality authors must classify risk for warning boxes-authoring tool For each warning box, we need to record:

- Event trying to avoid
- Likelihood
- Severity

The clinical expert also classified each warning in terms of the following:

- Swift action
- Stop hasty/over zealous action/drug interactions /point to which of two or more paths in guideline to take
- Referral
- When not to use this guideline/ Use appropriate guideline
- Order of action/ use appropriate equipment/ correct doses calculations
- Important statement applying whenever guideline is implemented
- Inappropriate

However, after the clinical expert had classified 380 warnings in terms of appropriateness, risk, and adding further comments in terms of context and analysis, it was determined that warning classification was extremely complex and would require dedicated research in terms of repeated studies and further analysis, all of which were outside the scope of this study. Warnings were therefore classified in terms of severity (Red, Amber and Blue).

Table 4: *Example Card Sorting cards provided to the participant for single criterion card sorting.*

5	Date (day, month, year) and time (using 24 hr clock) each entry, sign it, print your name and GMC number legibly with a contact bleep number or, if no bleep, telephone number and your grade	*Medical Records*
12	Expressed consent must be recorded in patient's clinical records	*Consent*
26	If any lesions or recurrent skin infections, or if any decontamination product causes skin irritation, contact occupational health	*Hand Hygiene*
39	IV adrenaline is hazardous, use only with extreme care, and under critical care supervision, for those in profound shock that is immediately life-threatening	*Acute Anaphylaxis*
45	Immediate treatment and investigations must run simultaneously	*Hypotention*
51	In the elderly, confusion can occur as the only symptom of meningitis in the absence of meningism or even of fever	*Community-Aquired Meningitis*
69	If Gram-negative bacilli grown in blood of patient returning from a typhoid endemic area (e.g. Indian sub-continent), give ceftriaxone 2 g IV by infusion daily; do not use ciprofloxacin as many strains of Salmonella typhi are resistant.	*Fever in a returning traveller*
85	Administer insulin and glucose infusions via same cannula using anti-siphon and anti- reflux valves (e.g. Vygon Protect-A-Line 2 extension set) through a large peripheral vein or central line – see Administration of IV insulin infusions and fluid infusions guideline	*Diabetic Ketoacidosis and Hyperosmolar Hyperglycaemic state.*
87	Blood glucose may rise as a result. Do not revert to sodium chloride 0.9%	*Diabetic Ketoacidosis and Hyperosmolar Hyperglycaemic state.*
91	Further information available from clinical biochemistry or from renal or endocrine teams	*Electrolyte Disturbances*

Table 5: *Card sort 1 - 4, categories created.*

Card Sort 1 Categories
High Risk
Low Risk
Card Sort 2 Categories
High Risk
Intermediate Risk
Low Risk
Card Sort 3 Categories
Background/Info for Patient Management
Referal
Admin
General Guidance
Card Sort 4 Categories
Management
Investigations
Diagnosis
Drug
Info
Biology (Bio)

Table 6: *Card sort 1 categories and cards.*

High Risk	**Low Risk**
12	45
85	287
227	91
39	87
210	51
184	223
347	96
235	26
160	5
179	89

Table 7: *Card sort 2 categories and cards.*

High Risk	**Intermediate Risk**	**Low Risk**
85	12	91
227	45	87
39	287	223
210	51	96
184	69	5
179	347	
	235	
	160	
	26	

3.5. Comparing the BCG app with current recommendations

In terms of the Royal College of Physicians (RCP) Health Informatics Unit checklist, the majority of the checklist criteria is met by the BCG app. This is especially true for sections 1 and 2 (The app functionality/development and usability). Figure 4 shows the comparison. In terms of section 3 there are several criteria that have not been met as these require further study. Measuring the impact on aspect such as clinical knowledge, actions and patient outcomes requires longitudinal studies for accurate and relevant data. However, it is worth note that the BCGs have been in use for over 20 years and as the app utilises the BCGs, it could be summised that these criteria will be met without issue once the app has been piloted and released. The app should also be approved for use as a Class 1 Medical Device, as per the MHRA UKCA mark criteria (formally CE mark) (https://www.gov.uk/guidance/regulating-medical-devices-in-the-uk).

Table 8: *Card sort 3 categories and cards.*

Background	**Referral**	**Admin**	**General Guidance**
223	235	12	187
210	347	5	85
69	26		45
287	96		160
51	91		
227	184		
179	39		

Table 9: *Card sort 4 categories and cards.*

Management	184	12	85	45
Investigations				
Diagnosis	91	69		
Drug	85			
Info	12	91		
Bio (Biology)	87			

Table 10: *Clinical warning hierarchy classification based on risk.*

V. High	**High**	**Int**	**Low**
5	25	43	20

4. DISCUSSION

The results present the impact of applying recommendations discussed in previous research publications (Mitchell et al. 2021). It is evident that the recommendations have enabled more succinct guidelines to be authored. The overall word count reduction of 23% evidences the reduction and therefore a more efficient set of guidelines (in line with efficiency proven through usability testing). The reduction in word count, tables and flowcharts contributes to a more usable system on mobile (smartphone)

Criteria	Criteria Met	Evidence
Who developed the app, and what's inside it?		
a) Is it clear who this app is for and how it should be used?	Y	Observations, Survey, User Testing
b) Is it clear which problem the app is designed to alleviate or what outcome it helps to promote?	Y	UCD Methods, previous use of the Bedside Clinical Guidelines
c) Do the app developer and sponsor seem well informed about this problem or outcome, and likely to be unbiased in their approach to it?	Y	UCD Methods, previous use of the Bedside Clinical Guidelines, Working with a clinical expert
d) Have they located sound, relevant, up-to-date evidence, images, video etc to use in their app?	Y	Updated by the Clinical Guidelines Librarian and provided in app
e) Do the app screens look well designed, is text clear?	Y	Designed with best practice and iterations based on feedback
f) Is it clear what data the app needs from the user with units defined, out of range detection and a 'clear last patient' button?	Y	Units are clear with examples, clear boundaries have been set in terms of calculation data, and each calculation tool or decision algorithm has a reset button
g) Does the app collect any identifiable patient information?	N	No identifiable patient data is collected
h) Does it seem to keep user and patient data secure and private?	N/A	-
i) If the app is designed to support any medical task,* is it CE marked?	N/A	This is currently in progress
2. How well does the app work?		
a) Is the app fast and easy to use in clinical settings?	Y	Using clinical scenarios during the think-aloud sessions evidenced that users can accurately and rapidly assimilate the information
b) Does the app give the user usable answers or advice quickly?	Y	As above
c) Do the answers, advice or calculated risks appear to be correct?	Y	Information has been validated working with a clinical expert
d) Is there a way to feedback user comments to the app developer?	Y	This will be provided in app as part of the next stage of pilot testing
3. Is there any evidence that the app does actually alleviate the problem?		
a) Have any studies been carried out to measure the impact of using the app on clinical or patient knowledge, actions or (preferably) patient outcomes?	N	This is planned for the next stages of the study
b) Were these studies independently conducted, well designed, large enough, and applicable to the user?	N/A	-
c) Did any study also examine health resource use, potential harms caused by the app, or quantify cost effectiveness?	N	This is planned for the next stages of the study
d) Overall, do the benefits of using this app seem likely to outweigh inconvenience and costs to the user?	Y	There is very little cost to the user as the BCGs are already utilised as part of clinical workflow
e) Is there any specific clinical scenario or patient subgroup in which using the app seems particularly likely to be useful?	Y	Bedside information retrieval

Figure 4: *App Checklist comparison*

devices. In some cases (see Table 1), the word count increased. This was due to several factors, the amalgamation of tables and flowcharts into the main text or where guidelines have been merged. Another

factor of increased word count is the repetition of warning text within each guideline. Although, the results do highlight the number of warnings contained within each guideline were reduced significantly. Reading rate was considered as a method of measuring efficiency. However, with limited access to clinicians and a dichotomy between readability for the general population and experts, it would be extremely difficult to accurately measure reading rate for each guideline - this is evidenced in a study by Bruce et al. (Bruce et al. 1981). The reading ease of the guidelines has an average score of 29.5 on the Flesch reading ease scale (Flesch 1948), meaning complex language is used (as expected) and therefore would be difficult to test without clinical knowledge. Warnings are clearly complex in terms of simplifying their presentation through categorisation and risk factors. It is difficult to determine how to present such information from the results obtained during the card sorting study. It does however highlight the need in working with a clinical expert. The card sorting pilot highlighted the complexity of clinical warnings at an early stage and therefore reduced the risk of conducting research that would not have been beneficial with other clinicians who have extremely limited access.

5. SUMMARY

This paper has highlighted the impact recommendations have had on the content of the BCGs. The application of the recommendations has had a significant impact on how the BCGs are authored and presented. By reducing the word count, tables and manual decision algorithms (flowcharts), the guidelines conform to a set of usability expectations/recommendations that benefit clinical users. Evidence provided in usability testing throughout the development of the prototype BCG app highlights the high usability rating and therefore the positive impact of applying these recommendations. Also highlighted is the necessity of working with a clinical expert to ensure the medical information contained within the BCG app remains valid and can be verified through simple checking processes. This study suggests that validation of information should be completed at all stages. At present BCG clinical guidelines do not have a validation process after the initial authoring and conversion for use in mobile device applications will require changes, as discussed, in the process of how they are designed, validated, and verified. The guidelines also need to present evidence to allow for an overview of how they are created. This can contribute to trust and learning. The outcome of the card sort and verification process highlights the complexity of clinical information, both in terms of knowledge requirement and methods of categorising the information. This research presents the benefits of building a close working relationship with an expert and utilising them throughout the study to ensure accurate and validated information. Working with a clinical expert was a crucial factor in recognising the importance and complexity of the information contained within the BCGs and highlights that any software application development process for clinical use must utilise clinical expertise. These processes have culminated in the production of a BCG app that meets the necessary criteria of trust and evidence in terms of development and information, as highlighted by the adherence to the RCP checklist.

REFERENCES

G Anthony Gorry, Howard Silverman, and Stephen G Pauker. Capturing clinical expertise: A computer program that considers clinical responses to digitalis. *The American journal of medicine*, 64(3): 452–460, 1978.

Jeremy Wyatt and David Spiegelhalter. Field trials of medical decision-aids: potential problems and solutions. In *Proceedings of the annual symposium on computer application in medical care*, page 3. American Medical Informatics Association, 1991.

Robert K Yin. *Case study research: Design and methods*, volume 5. sage, 2009.

Fariza Hanis Abdul Razak and Alan Dix. Doing the single person study in human computer interaction (hci). *Malaysian Journal of Computing (MJoC)*, 1 (1):1–14, 2010.

James Mitchell, Ed de Quincey, Charles Pantin, and Naveed Mustfa. The development of a point of care clinical guidelines mobile application following a user-centred design approach. In *International Conference on Human-Computer Interaction*, pages 294–313. Springer, 2020.

James Mitchell, EJ De Quincey, Charles Pantin, Naveed Mustfa, et al. 15 usability recommendations for delivering clinical guidelines on mobile devices. In *34th British HCI Conference*, pages 82–93. BCS Learning & Development, 2021.

Charles Pantin, John Mucklow, David Rogers, Marion Cross, Janine Wall, and The Bedside Clinical Guidelines Partnership. Bedside clinical guidelines: the missing link. *Clinical medicine*, 6 (1):98, 2006.

Jakob Nielsen and Thomas K Landauer. A mathematical model of the finding of usability problems. In *Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems*, pages 206–213, 1993.

Martin Maguire. Methods to support human-centred design. *International journal of human-computer studies*, 55(4):587–634, 2001.

David Brumley, M Ashby, Heather Robinson, and J Fisher. Improving access to clinical information in after hours community palliative care. *Australian Journal of Advanced Nursing, The*, 24(1):27–32, 2006.

Gordon Rugg and Peter McGeorge. The sorting techniques: a tutorial paper on card sorts, picture sorts and item sorts. *Expert Systems*, 14(2):80–93, 1997.

C Lee Ventola. Mobile devices and apps for health care professionals: uses and benefits. *Pharmacy and Therapeutics*, 39(5):356, 2014.

Paul Shekelle, Martin P Eccles, Jeremy M Grimshaw, and Steven H Woolf. When should clinical guidelines be updated? *Bmj*, 323(7305):155–157, 2001.

Thomas Lorchan Lewis and Jeremy C Wyatt. mhealth and mobile medical apps: a framework to assess risk and promote safer use. *Journal of medical Internet research*, 16(9):e3133, 2014.

Jeremy C Wyatt, Harold Thimbleby, Paul Rastall, Jan Hoogewerf, Darren Wooldridge, and John Williams. What makes a good clinical app? introducing the rcp health informatics unit checklist. *Clinical Medicine*, 15(6):519, 2015.

Bertram Bruce, Andee Rubin, and Kathleen Starr. Why readability formulas fail. *IEEE Transactions on Professional Communication*, (1):50–52, 1981.

R Flesch. A new readability yardstick journal of applied psychology 32: 221–233, 1948.