FACULTY OF MEDICINE AND HEALTH SCIENCES

USABILITY OF REGISTERED ELECTRONIC NURSING RECORDS USED BY HOME CARE NURSES IN FLANDERS

Word count: 5513

Klara Geltmeyer
Student number: 01010561

Promotor: Prof. dr. Pascal Coorevits

Master's thesis submitted for the master of science in nursing and midwifery

Academic year: 2016 – 2017

GHENT UNIVERSITY

Inhoudsopgave

Preface

I would like to thank my promotor prof. dr. Pascal Coorevits for his excellent guidance and support during this process. I also wish to thank all of the software vendors and respondents who participated in the study, without whose cooperation I would not have been able to conduct this research.

A special thanks goes out to my parents, who always support me in everything I do. I would also like to thank my sister, Jozefien, for helping and advising me while writing this thesis. I would also like to thank Jasper for having patience and supporting me through the process of this thesis.

A last thanks goes out to Dirk Musschoot for helping me writing this thesis with his expert knowledge of the English language.

At the end, I would like to thank you reader, for showing interest in this study.

Word count: 5.513

Abstract: English

Objective: We evaluated the usability of registered electronic nursing records used by home care nurses in Flanders.

Methodology: A cross-sectional study was designed using an online survey, including the System Usability Scale (SUS), the Questionnaire for User Interaction (QUIS) and a self-developed sociodemographic questionnaire. Nurses were recruited based on a randomised sample. The software vendors were used for recruitment of the nurses, to secure anonymity of the clients. An additional convenience sample was recruited via social media due to low response through the software vendors.

Results: Both questionnaires used were reliable with Cronbach's alpha's of 0.81 (SUS) and 0.98 (QUIS). The median [interquartile interval] SUS score was 67.50 [57.50 – 85.00] and the median [interquartile interval] QUIS score was 6.50 [5.50 – 7.50]. Ratings for all QUIS items were higher than neutral (5). A positive correlation (r_s = 0.608 (p=0.000)) was found between the SUS and the QUIS. A few significant (p<0,05) but low correlations were found between three sociodemographic variables (age, work experience in years and number of months using software) and QUIS items. Significant differences were found for the SUS score, the interface factor 'system capabilities', and several QUIS items by type of work (only nursing work vs. both administrative and nursing work).

Conclusion: Although this was an exploratory study the first results were encouraging, showing an ok to good usability of registered electronic nursing records used by home care nurses in Flanders.

Abstract: Nederlands

Doel: De gebruiksvriendelijkheid van gehomologeerde softwarepakketten gebruikt door thuisverpleegkundigen in Vlaanderen nagaan.

Methodologie: Een cross-sectionele studie werd ontwikkeld gebruikmakend van een online vragenlijst met de System Usability Scale (SUS), de Questionnaire for User Interaction Satisfaction (QUIS) en een zelfontwikkelde socio-demografische vragenlijst. Verpleegkundigen werden gerekruteerd op basis van een gerandomiseerde steekproef. Er werd samengewerkt met de software verkopers om de anonimiteit van de cliënten te verzekeren. Een bijkomende gelegenheidssteekproef werd gerekruteerd via sociale media omwille van een lage respons via de software verkopers.

Resultaten: Beide vragenlijsten waren betrouwbaar met een Cronbach's alpha van 0.81 (SUS) en 0.98 (QUIS). De mediaan [interkwartiel interval] SUS score was 67.50 [57.50 – 85.00] en de mediaan [interkwartiel interval] QUIS score was 6.50 [5.50 – 7.50]. Scores voor alle QUIS items waren hoger dan neutraal (5). Een positieve correlatie (r_s = 0.608 (p=0.000)) werd gevonden tussen de SUS en de QUIS. Enkele significante, maar lage correlaties (p<0,05) werden gevonden tussen drie socio-demografische variabelen (leeftijd, werkervaring in jaren en aantal maanden softwaregebruik) en QUIS items. Significante verschillen werden gevonden voor de SUS score, de factor 'mogelijkheden van het systeem' en verschillende QUIS items op basis van het type werk (enkel verpleegkundig werk t.o.v. zowel administratief als verpleegkundig werk).

Conclusie: De eerste resultaten van dit verkennend onderzoek zijn bemoedigend. Ze tonen een gebruiksvriendelijkheid van gehomologeerde softwarepakketten gebruikt door thuisverpleegkundigen in Vlaanderen aan die oké tot goed is.

The master thesis is written in article form. Extensive reporting of the systematic literature study is not a part of the written article. The literature study was previously assessed in the course 'Literature study'.

1 Introduction

The introduction of information technology (IT) in the health care industry has led to a number of changes. One of the most impactful has been the replacement of paper-based medical records by electronic medical records (Zahabi et al. 2015, Topaz et al. 2016). The European Institute for Health Records (EuroRec) defines an electronic medical record (EMR) as computerized legal clinical records created in care delivery organisations, such as hospitals and physician offices. Electronic health records (EHRs) make it easy to share medical information among stakeholders and to allow all professionally involved to follow the patient through various modalities of care from different care delivery organisations (Garets and Davis 2017). The term 'electronic nursing record (ENR)' is also used but no definition was available. The terms EMR, EHR and ENR are used simultaneously in literature.

In 2014, 96% of non-federal acute care hospitals in the U.S. possessed a certified EHR technology, of which 84% of hospitals in the U.S. adopted at least a basic EHR system, representing a nine-fold increase in EHR adoption since 2008 (Charles et al. 2015). The rate of EHR adoption by physicians in Europe are: The Netherlands (99%), Norway (97%), United Kingdom (96%), Germany (72%) and France (68%) (Schoen et al. 2012, Cho et al. 2016). There are no rates of EHR, EMR or ENR adoption available in Belgium.

Health information technology (HIT), including EMR, EHR and ENR, has great potential to enhance efficiency, improve quality of life and strengthen innovativeness in health and social care (Scandurra and Liljequist 2016). It is a key component of a comprehensive strategy to improve healthcare quality and patient safety (Ellsworth et al. 2017). It makes it possible for health care professionals to more effectively manage patient care through secure use and

sharing of health information (Page and Schadler 2014). A systematic review of Zahabi et al. (2015) shows that the use of EMRs has led to several improvements in health care such as increased information exchange among providers, increased quality of health care systems, decreased errors by health care workers and savings of time and money.

But as the adoption of EHRs by hospitals and ambulatory health care facilities continues to increase rapidly, dissatisfaction with EHRs continues to increase as well (Ratwani et al. 2015, Scandurra and Liljequist 2016, Topaz et al. 2016). Known areas of weakness and confusion are HIT usability, interoperability, and quality (Riskin et al. 2015). One of the ways organisations become aware of usability is trough critical incidents or “wake-up calls”. These could be patient safety issues linked to EHR usability or device purchases that greatly impact clinician’s productivity (Staggers and Rodney 2012). According to findings of physicians’ satisfaction surveys from EHR users in the U.S., more than 30% of respondents would not recommend their EHR to others, and nearly 50% of respondents were not satisfied with many of the largest available EHR systems (Cho et al. 2016). Not only clinicians, but also nurses, report low satisfaction with EHRs. The average satisfaction with the current state of EHRs used by nurses is reported as 4.5 (SD=2.3) on a scale from 1 (not at all satisfied) to 10 (very satisfied) (Topaz et al. 2016).

The International Standards Organization defines usability as “the degree to which a product or system can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use” (ISO 25000). It breaks the measurement of usability down into three separate components that must be defined in connection with the context of use:

- effectiveness: whether people can actually complete their tasks and achieve their goals
- efficiency: the extent to which they expend resource in achieving their goals

- satisfaction: the level of comfort they experience in achieving those goals (Brooke 2013).

Implementing EHRs without a focus on usability is the largest barrier to widespread adoption of it (Page and Schadler 2014, Riskin et al. 2015). Usability can be a strategic asset, saving organisations revenue, preventing errors and improving the users experience with products ranging from EHRs to medical devices (Staggers and Rodney 2012). It is required for HIT to be effective and it is foundational to deriving benefit from software (Riskin et al. 2015). Despite the fact that usability is a major issue for health information technology adoption, little is known about the usability of EHRs (Cho et al. 2016).

As stated above usability of many EHRs continues to be suboptimal despite the widespread adoption of EHRs. Some vendors fail to meet usability standards, resulting in clinician frustration and patient safety hazards (Ratwani et al. 2017). Recognising the importance of usability in the US, the federal Office of the National Coordinator for Health Information Technology (ONC) has established user-centered design requirements that must be met before vendors of EHRs receive certification (Ratwani et al. 2017, Ellsworth et al. 2017). However, an analysis of reports suggests that many vendors are not adhering to the certification requirements (Ratwani et al. 2017). In Belgium a homologation process of software took place in home care nursing. This resulted in a list of criteria for the registration of software of electronic nursing records. After conformity tests from 2013 till 2015, 19 software packages passed all criteria, of which 13 software packages in Flanders (2017).

A systematic review shows that the majority of the studies on usability of EHRs were conducted in the United States (53%) and the Netherlands (8%), and no other country had more than four articles (Ellsworth et al. 2017). Nurses' perceptions on the usability of EHRs are rarely surveyed and remain therefore largely unknown (Topaz et al. 2016). A systematic review conducted between 2003 and 2009 shows that only 7 out of 346 identified usability studies on HIT examined systems used by nurses (Yen and Bakken 2012). To the best of our

knowledge, no Belgian research has been carried out on the usability of EMRs, EHRs or ENRs.

Most of the usability studies are conducted in a hospital setting, only one study was found evaluating the use of EHRs in home care nursing and named itself ‘the first know evaluation of EHRs in home care’. Reporting usability issues, such as slow screen changes, difficulty logging on and essential clinical information not displayed in the documentation window, as one of the observed barriers to use EHRs (Sockolow et al. 2012).

To ensure the best utilisation of health IT, including EHRs, it is essential to pay attention to usability, keeping in mind its intended users, task and environment (Yen and Bakken 2012). Therefore, this study will investigate the usability of registered electronic nursing records used by home care nurses in Flanders. Focus will be given to evaluating the whole system and subparts of the software, such as the screen, terminology etc. This research can have major importance in optimising the usability of EHRs and it may thus aid in the widespread implementation of a user friendly EHR system, contributing to an enhanced efficiency in health and social care.

2 Methodology

A cross-sectional study was designed to evaluate the usability of registered electronic nursing records by home care nurses in Flanders.

2.1 Survey

The study incorporated two usability questionnaires: the System Usability Scale (SUS) and the Questionnaire for User Interaction Satisfaction (QUIS). Both are validated instruments (Brooke 1996, Norman and Schneiderman 2017). There is currently little guidance on how to perform usability evaluations. A systematic review shows that the most frequent methods used are surveys (37%) and think-aloud methods (19%), which combined, account for more than half of all the usability evaluations reviewed. The most common existing and validated survey used is the SUS (20%), followed closely by the QUIS (16%) (Ellsworth et al. 2017). The SUS, QUIS and several socio-demographic questions were incorporated in one online survey (appendix 1). This survey was added to an online survey tool 'Limesurvey®' containing three parts in the following order: SUS, QUIS and socio-demographic questions.

2.1.1 System Usability Scale

The SUS is a simple, ten-item scale giving a global view of subjective assessments of usability. The SUS uses alternating items to avoid response biases, by alternating positive and negative statements the goal is to have respondents read each statement and make an effort to think whether they agree or disagree with it (Brooke 2013). The items are rated on a 5-point scale rating from 1 (I strongly agree) to 3 (I agree nor disagree) to 5 (I strongly disagree) (Bangor et al. 2009, Brooke 1996, Sturm et al. 2002). It is called a 'quick and dirty' usability scale, referring to its easiness of use (Brooke 2013).

Research indicates that the SUS score might be biased in a remote, unsupervised, international usability evaluation due to a misunderstanding or misinterpretation of terminology in the scale, in particular item 8 ('I found the system very cumbersome to use'). There could be opted to prevent such issues

by rewording item 8 in ‘I found the system very awkward to use’, which clarifies the wording while retaining the original language (Finstad 2006, Bangor et al. 2009, Lewis and Sauro 2009). But in this research there has been opted to translate the SUS to Dutch according to previous research using a Dutch SUS questionnaire (Jansen-Kosterink et al. 2012). Unofficial translations into Spanish, French and Dutch have been reported and research shows that, working with a Dutch research team, a translated version had similar internal reliability to the original English version (Brooke 2013). The internal reliability will also be tested in this research. SUS was made freely available by the creator John Brooke.

To summarise the reasons for choosing the SUS on the basis of several characteristics that make its use attractive:

- it is composed of only ten statements
- it is non-proprietary
- it is technology agnostic
- the result of the SUS is one single score (Bangor et al. 2009).

The SUS yields a single number representing a composite measure of the overall usability of the software, individual items are not meaningful on their own (Brooke 1996). The SUS score is calculated by summing the score contributions of each item. For items 1, 3, 5, 7 and 9 the score contribution is the scale position minus 1. For items 2, 4, 6, 8 and 10 the score contribution is five minus the scale position. To obtain the overall value of the SUS score the sum of the scores has to be multiplied by 2.5. SUS scores range from 0 to 100 (Brooke 1996).

2.1.2 Questionnaire for User Interaction Satisfaction

The QUIS is a tool developed at the University of Maryland. A student license for the current version QUIS 7.0 was obtained. We chose to use the long version of the QUIS. The QUIS contains a demographic questionnaire, a measure of overall system satisfaction and measures of nine specific interface factors (screen factors, terminology and system feedback, learning factors, system capabilities, technical manuals, online tutorials, multimedia, teleconferencing, and software installation). Each factor measures the users' overall satisfaction with facets of the interface on a 9-point scale. The questionnaire is designed to be configured by including only the sections that are of the user's interest (Norman and Schneiderman 2017). In this study the demographic questionnaire will not be included because there was opted to include a self-developed questionnaire. The measure of overall user reactions and four of the nine specific interface factors will be included, screen factors, terminology and system feedback, learning factors and system capabilities. Other studies evaluating the usability of an EHR or EMR made the same inclusion decision concerning the specific interface factors (Hoyt et al. 2013, Gadd et al. 2011, Jaspers et al. 2008).

2.2 Vendor and sample selection

Participants were recruited based on a randomised sample. First, the vendors of the 13 registered software packages in Flanders were asked to participate via e-mail. Upon confirmation they were sent an e-mail with a list of 50 randomised client numbers and a standard mail containing the information letter for the clients and a link to the online survey. They were asked to contact the clients using the standard mail to assure anonymity of the clients. A reminder e-mail was sent to all participants of the participating vendors. An additional convenience sample was recruited via social media due to low response via the vendors of the software packages. Data collection took place from 29/03/2017 until 08/05/2017.

2.3 Informed consent

The participants were contacted through e-mail by the software vendors. The participants were informed trough a standard mail containing all the necessary information. Before starting the online survey, the participants were shown the information letter again and were asked for their online consent to participate in the study. This study was approved by the ethics committee of the University Hospital of Ghent.

2.4 Statistical analysis

Statistical analyses were performed with SPSS version 24.0. Based on interpretation of histograms, QQ-plots and Shapiro-Wilk tests on all the variables, non-parametric tests were performed.

A SUS score was calculated in the manner described above. In the online version of the SUS there was a ‘no answer’ option. Those ‘no answer’ responses were changed to a score of ‘3’ (Bangor et al. 2009). Three different QUIS scores will be shown: overall median QUIS score, QUIS score per interface factor and QUIS score per item. In the obtained paper version of the QUIS there was a ‘not applicable (NA)’ option in addition to the rating scale from 1 to 9. The NA option was also included in the online version but during data analysis NA responses were treated as missing variables. Descriptive statistics of SUS and QUIS scores were reported using medians and interquartile intervals.

Spearman correlation coefficients were calculated between SUS and the first QUIS factor, between SUS and the overall median QUIS score, between SUS and three sociodemographic variables (age, work experience in years and number of months using software), between the QUIS interface factors and the three abovementioned variables, and finally between the QUIS items and the same three sociodemographic variables.

In order to investigate possible differences in SUS and QUIS scores between type of work (only nursing work or both administrative and nursing work), Mann-

Whitney U tests were performed. For the correlation coefficients and the Mann-Whitney U tests the significance level was set at $\alpha = 0,05$.

Cronbach's alpha's were calculated for both the SUS and the QUIS to test the reliability of both instruments.

3 Results

3.1 Socio-demographic data

Six out of 13 homologated software packages in Flanders were included in the research, giving us a total of 300 nurses that were asked to complete the questionnaire, of which only 37 completed it, resulting in a response rate of 12%.

An additional convenience sample of 20 participants was recruited via social media. Resulting in a total of 57 nurses who completed the questionnaire, of which 45 completed the whole questionnaire including the socio-demographic questions (last part of the questionnaire).

Nurses ranging from the age of 23-66 completed the questionnaire, with a median age [interquartile interval] of 45 years [32.50-53]. The research population consisted of four male and 41 female participants.

The survey was conducted on a relatively experienced workforce, with half of the software users using their software package 24 months or longer ([15,48]) and 16 of them having previous experience with other software packages. 19 of the participants were self-employed nurses and the other 26 were employed in an organisation. Half (n=23) of the nurses only conducted nursing work with the software and the other half (n=22) conducted both administrative and nursing work. The nurses had following education (highest degree): HBO-5 (n=23), bachelor in nursing (n=19), bachelor after bachelor (n=1), master degree (n=1), master after master (n=1). The working experience varied between 1 and 37 years. Half of the (n=22) nurses worked already longer than 20 years, and the other half (n=23) worked less than 20 years.

3.2 Surveys

A Cronbach's alpha was calculated for both the SUS and the QUIS to test the reliability. The SUS had a Cronbach's alpha of 0.81 (n=57) and the QUIS had a Cronbach's alpha of 0.98 (n=20).

The median [interquartile interval] SUS score in this research was 67.50 [57.50 - 85.00].

The overall median [interquartile interval] of the QUIS was 6.50 [5.50 - 7.50] and ratings for all usability items were greater than neutral (5). Table 1 shows the five specific interface factors of the QUIS with their subscale items on the basis of their median and interquartile range.

Table 1: QUIS interface factors and items with their median and interquartile range

	QUIS items	**Scale: 1**	**Median [Interquartile interval]**	**N**	**Scale: 9**	
Overall User Reactions	The software is	Terrible	6.50 [6.00 – 7.25]	54	Wonderful	Median = 6.50 [5.50 – 7.00]
	The software is	Frustrating	6.00 [5.00 – 7.00]	55	Satisfying	
	The software is	Dull	6.00 [5.00 – 7.50]	53	Stimulating	
	The software is	Difficult	7.00 [6.00 – 7.00]	55	Easy	
	The software has	Inadequate power	6.00 [5.00 – 8.00]	54	Adequate power	
	The software is	Rigid	6.00 [5.00 – 7.00]	53	Flexible	
Screen	Screen layouts were helpful	Never	7.00 [6.00 – 7.50]	49	Always	Median = 7.00 [6.00 – 8.00]
	Amount of information that can be displayed on screen	Inadequate	7.00 [6.00 – 8.00]	52	Adequate	
	Arrangement of information on screen	Illogical	7.00 [6.00 – 8.00]	51	Logical	
	Sequence of screens	Confusing	7.00 [6.00 – 8.00]	52	Clear	
	Next screen in a sequence	Unpredictable	7.00 [6.00 – 8.00]	50	Predictable	
	Going back to the previous screen	Impossible	7.00 [6.00 – 8.00]	52	Easy	
	Progression of work related tasks	Confusing	7.00 [6.00 – 8.00]	52	Clearly marked	

Terminology and system information	Use of terminology throughout system	Inconsistent	7.00 [6.00 – 8.00]	49	Consistent	Median = 7.00 [5.50 – 8.00]
	Work related terminology	Inconsistent	8.00 [6.00 – 8.00]	49	Consistent	
	Computer terminology	Inconsistent	7.00 [5.00 – 8.00]	45	Consistent	
	Messages which appear on screen	Inconsistent	7.00 [6.00 – 8.00]	47	Consistent	
	Positions of instructions on the screen	Inconsistent	7.00 [6.00 – 8.00]	45	Consistent	
	Messages which appear on screen	Confusing	7.00 [6.00 – 9.00]	46	Clear	
	Instructions for commands or functions	Confusing	7.00 [6.00 – 8.00]	47	Clear	
	Instructions for correcting errors	Confusing	6.00 [5.00 – 8.00]	47	Clear	
	Computer keeps you informed about what it is doing	Never	6.00 [4.00 – 7.50]	45	Always	
	Animated cursors keep you informed	Never	5.00 [2.00 – 7.00]	34	Always	
	Performing an operation leads to a predictable result	Never	7.00 [6.00 – 8.00]	45	Always	
	Controlling amount of feedback	Impossible	6.00 [5.00 – 7.00]	43	Easy	
	Length of delay between operation	Unacceptable	5.00 [3.00 – 6.75]	44	Acceptable	
	Error messages	Unhelpful	6.00 [5.00 – 9.00]	46	Helpful	
Learning	Learning to operate the system	Difficult	7.00 [5.75 – 8.00]	50	Easy	Median = 6.50 [5.50 – 8.00]
	Getting started	Difficult	6.00 [5.75 – 8.00]	50	Easy	
	Learning advanced features	Difficult	6.00 [5.00 – 7.00]	48	Easy	
	Time to learn to use the system	Slow	6.00 [5.00 – 8.00]	48	Fast	
	Tasks can be performed in a straight-forward manner	Never	7.00 [6.00 – 8.00]	48	Always	
	Number of steps per task	Too many	6.50 [5.00 – 8.00]	48	Just right	
	Steps to complete a task follow a logical sequence	Never	7.00 [6.00 – 7.75]	48	Always	
	Feedback on the completion of steps	Unclear	6.00 [5.00 – 7.00]	45	Clear	

System Capabilities	System speed	Too slow	5.00 [3.00 – 7.75]	48	Fast enough	Median = 6.00 [4.625 – 8.00]
	Response time for most operations	Too slow	6.00 [4.00 – 8.00]	45	Fast enough	
	Rate information is displayed	Too slow	5.00 [4.00 – 8.00]	37	Fast enough	
	The system is reliable	Never	7.00 [6.00 – 9.00]	46	Always	
	Operations are	Undependable	7.00 [5.25 – 8.75]	40	Dependable	
	System failures occur	Frequently	5.00 [3.00 – 8.00]	47	Seldom	
	System warns you about potential problems	Never	5.00 [3.00 – 8.00]	43	Always	
	Correcting your mistakes	Difficult	5.00 [4.00 – 7.00]	43	Easy	
	Correcting typos	Complex	6.00 [4.75 – 8.25]	42	Simple	
	Ability to undo operations	Inadequate	6.00 [4.00 – 8.00]	43	Adequate	
	Ease of operation depends on your level of experience	Never	6.00 [4.00 – 8.00]	45	Always	
	You can accomplish tasks knowing only a few commands	With difficulty	7.00 [5.75 – 8.00]	46	Easily	
	You can use features/shortcuts	With difficulty	5.50 [3.00 – 7.00]	38	Easily	

The first interface factor is the ‘overall user reactions’. All of the subscale items have medians above 6. The next interface factor, ‘the screen’ has medians of the subscale items all equal to 7. The third interface factor ‘terminology and system information’ has one item with a median of 8 that sticks out for ‘work related terminology (inconsistent-consistent)’. Two subscale items have a neutral score of 5; ‘animated cursors keep you informed (never-always)’ and ‘length of delay between operations (unacceptable-acceptable)’. The fourth interface factor ‘learning’ has median scores between 6.00 and 7.00 on all items. The last interface factor ‘system capabilities’ has the lowest median scores. Four items have a neutral score of 5: ‘rate information is displayed (too slow-fast enough)’, ‘system failures occur (frequently-seldom)’, ‘system warns you about potential problems (never-always)’ and ‘correcting your mistakes (difficult-easy)’.

When comparing the median scores of the different interface factors, the lowest median score [interquartile interval] of 6.00 [4.625 – 8.00] represents the 'system capabilities'. Followed by two factors with a median score of 6.50, the 'overall user reactions' and 'learning'. The highest median scores of 7 of the interface factors are the 'screen' and 'terminology and system information'.

3.3 Correlations

3.3.1 Correlations between SUS score and the first QUIS interface factor

A low but significant correlation was found between the SUS score and the first QUIS interface factor 'overall user reactions' with $r_s = 0.433$ (p=0.001).

3.3.2 Correlations between SUS score and overall median QUIS score

A moderate significant correlation was found between the SUS score and the overall median of the QUIS score with $r_s = 0.608$ (p=0.000).

3.3.3 Correlations between SUS score and three sociodemographic variables

Table 2 presents the correlation coefficients between the SUS score and three sociodemographic variables (age, work experience in years and number of months using software). For each correlation, low and non-significant correlation coefficients (all p>0,05) were found.

Table 2: Correlations (p-values) between SUS score and three sociodemographic variables

	Age	**Work experience**	**Use software**
SUS	-0.146 (p=0.338)	-0.229 (p=0.130)	0.064 (p=0.700)

3.3.4 Correlations between QUIS factors and three sociodemographic variables

The correlations between the five interface factors of the QUIS and the three sociodemographic variables (age, work experience in years and number of months using software) are shown in table 3. Low and non-significant correlation coefficients were found in each case (all p>0,05).

Table 3: Correlations (p-values) between QUIS factors and three sociodemographic variables

QUIS factor	Age	Work experience	Time of use
Overall user reactions	0.085 (p=0.580)	0.044 (p=0.774)	-0.100 (p=0.544)
Screen	-0.059 (p=0.700)	-0.107 (p=0.483)	-0.205 (p=0.211)
Terminology and system information	0.041 (p=0.789)	0.010 (p=0.950)	-0.122 (p=0.458)
Learning	-0.031 (p=0.842)	-0.077 (p=0.617)	-0.091 (p=0.580)
System capabilities	0.174 (p=0.253)	0.140 (p=0.359)	-0.106 (p=0.519)

3.3.5 Correlations between QUIS items and three sociodemographic variables

Table 4: Correlations (p-values) between QUIS items and three sociodemographic variables

	QUIS items	Age	Work experience	Software use
Overall user reactions	Overall reactions to the software (terrible – wonderful)	0.059 (p=0.704)	0.017 (p=0.913)	-0.023 (p=0.889)
	Overall reactions to the software (frustrating – satisfying)	0.177 (p=0.245)	0.145 (p=0.341)	-0.048 (p=0.773)
	Overall reactions to the software (dull – stimulating)	0.273 (p=0.073)	0.229 (p=0.135)	-0.066 (p=0.694)
	Overall reactions to the software (difficult – easy)	-0.087 (p=0.568)	-0.129 (p=0.398)	-0.122 (p=0.458)
	Overall reactions to the software (inadequate power – adequate power)	0.008 (p=0.957)	-0.048 (p=0.755)	-0.165 (p=0.324)
	Overall reactions to the software (rigid – flexible)	0.071 (p=0.649)	0.005 (p=0.974)	-0.028 (p=0.866)
Screen	Screen layouts were helpful (never – always)	-0.128 (p=0.420)	-0.178 (p=0.260)	-0.243 (p=0.147)
	Amount of information that can be displayed on screen (inadequate – adequate)	-0.043 (p=0.777)	-0.087 (p=0.570)	-0.253 (p=0.120)
	Arrangement of information on screen (illogical – logical)	-0.140 (p=0.363)	-0.144 (p=0.350)	**-0.323 (p=0.048)**
	Sequence of screens (confusing – clear)	-0.069 (p=0.653)	-0.100 (p=0.515)	-0.173 (p=0.292)
	Next screen in a sequence (unpredictable - predictable)	-0.017 (p=0.913)	-0.035 (p=0.825)	-0.267 (p=0.110)
	Going back to the previous screen (impossible – easy)	-0.026 (p=0.864)	-0.052 (p=0.737)	-0.168 (p=0.307)
	Progression of work related tasks (confusing – clearly marked)	0.054 (p=0.727)	0.012 (p=0.937)	0.010 (p=0.950)
Terminology and system information	Use of terminology throughout the software (inconsistent – consistent)	0.069 (p=0.657)	0.047 (p=0.760)	0.032 (p=0.851)
	Work related terminology (inconsistent – consistent)	0.056 (p=0.717)	0.055 (p=0.724)	-0.015 (p=0.927
	Computer terminology (inconsistent – consistent)	0.058 (p=0.604)	0.050 (p=0.758)	0.038 (p=0.832)
	Messages which appear on screen (inconsistent – consistent)	0.071 (p=0.652)	0.040 (p=0.797)	-0.004 (p=0.982)
	Position of instructions on the screen (inconsistent – consistent)	0.195 (p=0.222)	0.146 (p=0.364)	0.074 (p=0.671)
	Messages which appear on screen (confusing – clear)	0.046 (p=0.774)	-0.010 (p=0.951)	-0.091 (p=0.597

	Instructions for commands or functions (confusing – clear)	-0.060 (p=0.704)	-0.092 (p=0.556)	-0.036 (p=0.831)
	Instructions for correcting errors (confusing – clear)	0.159 (p=0.314)	0.120 (p=0.449)	0.001 (p=0.997)
	Computer keeps you informed about what it is doing (never – always)	0.225 (p=0.157)	0.196 (p=0.220)	-0.088 (p=0.614)
	Animated cursors keep you informed (never – always)	0.311 (p=0.094)	0.261 (p=0.164)	0.070 (p=0.733)
	Performing an operation leads to a predictable result (never – always)	-0.016 (p=0.924)	-0.044 (p=0.788)	-0.160 (0.358)
	Controlling amount of feedback (impossible – easy)	0.225 (p=0.175)	0.135 (p=0.418)	0.060 (p=0.743)
	Length of delay between operation (unacceptable – acceptable)	0.158 (p=0.338)	0.093 (p=0.573)	0.124 (p=0.493)
	Error messages (unhelpful – helpful)	**0.415 (p=0.007)**	**0.355 (p=0.023)**	0.074 (p=0.672)
Learning	Learning to operate the system (difficult – easy)	-0.190 (p=0.212)	-0.243 (p=0.108)	-0.094 (p=0.571)
	Getting started (difficult – easy)	-0.020 (p=0.896)	-0.067 (p=0.663)	-0.054 (p=0.742)
	Learning advanced features (difficult – easy)	-0.138 (p=0.378)	-0.183 (p=0.240)	-0.180 (p=0.286)
	Time to learn to use the system (slow – fast)	-0.138 (p=0.379)	-0.171 (p=0.272)	-0.064 (p=0.707)
	Tasks can be performed in a straight-forward manner (never – always)	0.071 (p=0.651)	0.026 (p=0.868)	-0.099 (p=0.560)
	Number of steps per task (too many – just right)	0.147 (p=0.346)	0.116 (p=0.457)	-0.069 (p=0.684)
	Steps to complete a task follow a logical sequence (never – always)	0.102 (p=0.514)	0.045 (p=0.772)	-0.061 (p=0.718)
	Feedback on the completion of steps (unclear – clear)	0.160 (p=0.324)	0.116 (p=0.477)	-0.112 (p=0.530)
System capabilities	System speed (too slow – fast enough)	0.282 (p=0.061)	0.225 (p=0.137)	0.178 (p=0.278)
	Response time for most operations (too slow – fast enough)	0.234 (p=0.135)	0.184 (p=0.244)	0.093 (p=0.590)
	Rate information is displayed (too slow – fast enough)	0.311 (p=0.074)	0.260 (p=0.137)	0.096 (p=0.619)
	The system is reliable (never – always)	0.123 (p=0.430)	0.086 (p=0.583)	0.062 (p=0.716)
	Operations are (undependable – dependable)	0.158 (p=0.349)	0.093 (p=0.586)	0.070 (p=0.707)
	System failures occur (frequently – seldom)	0.126 (p=0.415)	0.095 (p=0.540)	-0.170 (p=0.308)
	System warns you about potential problems (never – always)	**0.413 (p=0.008)**	**0.355 (p=0.024)**	0.240 (p=0.172)
	Correcting your mistakes (difficult – easy)	-0.115 (p=0.480)	-0.121 (p=0.455)	**-0.383 (p=0.023)**
	Correcting typos (complex – simple)	0.296 (p=0.067)	0.276 (p=0.089)	0.009 (p=0.958)
	Ability to undo operations (inadequate – adequate)	0.221 (p=0.170)	0.219 (p=0.175)	**-0.346 (p=0.042)**
	Ease of operation depends on your level of experience (never – always)	0.346 (p=0.025)	0.263 (p=0.092)	0.074 (p=0.668)

	You can accomplish tasks knowing only a few commands (with difficulty – easily)	-0.180 (p=0.248)	-0.202 (p=0.195)	-0.310 (p=0.062)
	You can use features/shortcuts (with difficulty – easily)	-0.183 (p=0.293)	-0.223 (p=0.198)	**-0.430 (p=0.018)**

In general, the vast majority of the correlation coefficients demonstrated to be non-significant (p>0,05). However, several low but significant correlations were found. The sociodemographic variable 'age' was positively correlated with two QUIS items: 'error messages (unhelpful-helpful)' with r_s =0.415 (p=0.007) and 'system warns you about potential problems (never-always)' with r_s =0.413 (p=0.008). The second sociodemographic variable 'work experience in years' was positively correlated with the same two QUIS items: 'error messages (unhelpful-helpful)' with r_s =0.355 (p=0.023) and 'system warns you about potential problems (never-always)' with r_s =0.355 (p=0.024). And the last sociodemographic variable (number of months using software) was negatively correlated with four QUIS items: 'arrangement of information on screen (illogical-logical)' with r_s =-0.323 (p=0.048), 'correcting your mistakes (difficult-easy)' with r_s =-0.383 (p=0.023), 'ability to undo operations (inadequate-adequate)' with r_s =-0.346 (p=0.042) and 'you can use features/shortcuts (with difficulty-easily)' with r_s =-0,430 (p=0.018).

3.4 Differences in SUS and QUIS scores between type of work

Table 5 shows the SUS and QUIS scores with the differences between type of work, representing only nursing work or both nursing and administrative work.

Table 5: Median [interquartile intervals] SUS and QUIS scores between type of work

	Only nursing work	**Nursing + administrative work**	**p-value**
SUS score			
	62.5 [55 – 80]	77.5 [61.875 – 87.5]	**0.048**
QUIS overall score			
	6.00 [5.00 – 7.00]	7.00 [6.00 – 8.00]	0.080
QUIS factors			
Overall user reactions	6.50 [5.50 – 7.00]	6.50 [6.00 – 8.00]	0.396
Screen	7.00 [6.00 – 8.00]	7.00 [6.00 – 8.00]	0.276
Terminology and system information	7.00 [5.00 – 8.00]	7.00 [6.00 – 8.00]	0.317
Learning	6.50 [5.50 – 7.50]	7.00 [5.88 – 8.00]	0.696
System capabilities	6.00 [4.00 – 7.00]	7.00 [6.00 – 8.00]	**0.038**

QUIS items				
Overall user reactions	Overall reactions to the software (terrible – wonderful)	6.50 [5.75 – 7.25]	7.00 [6.00 – 8.00]	0.392
	Overall reactions to the software (frustrating – satisfying)	6.00 [5.00 – 7.00]	7.00 [5.75 – 8.00]	0.432
	Overall reactions to the software (dull – stimulating)	6.00 [5.75 – 7.00]	6.50 [5.00 – 8.00]	0.402
	Overall reactions to the software (difficult – easy)	7.00 [6.00 – 8.00]	7.00 [5.75 – 7.00]	0.546
	Overall reactions to the software (inadequate power – adequate power)	6.00 [5.00 – 7.25]	7.00 [5.00 – 8.00]	0.426
	Overall reactions to the software (rigid – flexible)	6.00 [5.00 – 7.00]	7.00 [5.50 – 8.00]	0.062
Screen	Screen layouts were helpful (never – always)	7.00 [6.00 – 7.50]	7.00 [6.00 – 8.00]	0.585
	Amount of information that can be displayed on screen (inadequate – adequate)	7.00 [6.00 – 8.00]	7.00 [6.00 – 8.00]	0.377
	Arrangement of information on screen (illogical – logical)	7.00 [6.00 – 8.00]	7.00 [6.75 – 8.00]	0.576
	Sequence of screens (confusing – clear)	7.00 [6.00 – 8.00]	7.00 [7.00 – 8.25]	0.117
	Next screen in a sequence (unpredictable - predictable)	7.00 [6.00 – 8.00]	7.00 [6.00 – 8.00]	0.356
	Going back to the previous screen (impossible – easy)	7.00 [6.00 – 9.00]	7.00 [6.00 – 8.25]	0.982
	Progression of work related tasks (confusing – clearly marked)	6.00 [6.00 – 8.00]	7.00 [6.00 – 8.00]	0.302
Terminology and system information	Use of terminology throughout the software (inconsistent – consistent)	7.00 [6.00 – 8.00]	8.00 [6.50 – 8.00]	0.647
	Work related terminology (inconsistent – consistent)	8.00 [6.00 – 9.00]	8.00 [7.00 – 8.50]	0.726
	Computer terminology (inconsistent – consistent)	6.00 [5.00 – 8.25]	7.00 [5.75 – 8.00]	0.455
	Messages which appear on screen (inconsistent – consistent)	7.00 [6.00 – 8.00]	7.00 [6.00 – 8.00]	0.663
	Position of instructions on the screen (inconsistent – consistent)	6.00 [5.00 – 7.50]	7.50 [7.00 – 8.00]	**0.011**
	Messages which appear on screen (confusing – clear)	7.00 [6.00 – 9.00]	7.00 [6.50 – 9.00]	0.907
	Instructions for commands or functions (confusing – clear)	6.00 [5.75 – 8.00]	7.00 [6.00 – 8.50]	0.275
	Instructions for correcting errors (confusing – clear)	6.00 [3.50 – 7.50]	7.00 [5.00 – 8.00]	0.237
	Computer keeps you informed about what it is doing (never – always)	6.00 [3.75 – 7.00]	7.00 [5.00 – 8.00]	0.199
	Animated cursors keep you informed (never – always)	3.50 [1.00 – 6.25]	6.00 [3.25 – 7.75]	0.081
	Performing an operation leads to a predictable result (never – always)	7.00 [5.25 – 8.00]	7.50 [6.25 – 8.00]	0.269

	Controlling amount of feedback (impossible – easy)	6.00 [4.50 – 7.00]	7.00 [5.00 – 7.75]	0.557
	Length of delay between operation (unacceptable – acceptable)	5.00 [2.50 – 6.00]	5.00 [3.75 – 9.00]	0.366
	Error messages (unhelpful – helpful)	6.00 [5.00 – 7.75]	7.00 [6.00 – 9.00]	0.067
Learning	Learning to operate the system (difficult – easy)	7.00 [6.00 – 9.00]	6.50 [4.75 – 8.00]	0.273
	Getting started (difficult – easy)	6.00 [5.00 – 8.00]	6.50 [6.00 – 8.00]	0.890
	Learning advanced features (difficult – easy)	6.00 [4.75 – 7.25]	6.00 [5.50 – 7.00]	0.729
	Time to learn to use the system (slow – fast)	6.50 [5.00 – 8.00]	6.00 [5.00 – 8.00]	0.629
	Tasks can be performed in a straight-forward manner (never – always)	7.00 [5.00 – 8.00]	7.00 [6.00 – 8.00]	0.372
	Number of steps per task (too many – just right)	6.00 [5.00 – 7.25]	7.00 [4.50 – 8.00]	0.514
	Steps to complete a task follow a logical sequence (never – always)	6.50 [6.00 – 7.00]	7.00 [5.50 – 8.00]	0.344
	Feedback on the completion of steps (unclear – clear)	6.00 [4.75 – 7.00]	7.00 [5.75 – 8.00]	0.111
System capabilities	System speed (too slow – fast enough)	5.00 [2.00 – 7.00]	7.00 [5.00 – 9.00]	**0.038**
	Response time for most operations (too slow – fast enough)	5.50 [3.75 – 7.00]	7.00 [5.25 – 9.00]	0.071
	Rate information is displayed (too slow – fast enough)	4.00 [2.50 – 6.50]	7.00 [5.00 – 9.00]	**0.012**
	The system is reliable (never – always)	7.00 [6.00 – 8.00]	8.00 [7.00 – 9.00]	0.186
	Operations are (undependable – dependable)	7.00 [5.00 – 8.75]	7.00 [7.00 – 9.00]	0.256
	System failures occur (frequently – seldom)	5.00 [3.00 – 7.00]	7.50 [4.50 – 9.00]	0.098
	System warns you about potential problems (never – always)	4.50 [1.00 – 6.00]	7.00 [4.25 – 8.75]	**0.015**
	Correcting your mistakes (difficult – easy)	5.00 [4.00 – 6.75]	6.50 [4.25 – 7.00]	0.171
	Correcting typos (complex – simple)	6.00 [2.50 – 8.00]	7.00 [5.75 – 9.00]	0.114
	Ability to undo operations (inadequate – adequate)	5.00 [3.00 – 8.00]	6.00 [5.50 – 8.00]	0.251
	Ease of operation depends on your level of experience (never – always)	5.00 [3.75 – 8.00]	7.00 [6.00 – 8.00]	0.193
	You can accomplish tasks knowing only a few commands (with difficulty – easily)	7.00 [5.50 – 8.00]	7.00 [6.00 – 8.00]	0.656
	You can use features/shortcuts (with difficulty – easily)	5.00 [3.00 – 8.00]	6.00 [3.50 – 7.00]	0.920

A first significant difference was found between the SUS and type of work (p=0.048). Subsequently a significant difference was found between the QUIS interface factor ‘system capabilities’ and type of work (p=0.038). There were also several significant differences found between type of work and the following QUIS items: ‘position of instructions on the screen (inconsistent-consistent)’ (p=0.011), ‘system speed (too slow-fast enough)’ (p=0.038), ‘rate information is displayed (too slow-fast enough)’ (p=0.012), ‘system warns you about potential problems (never-always)’ (p=0.015). All the significant differences show higher QUIS scores given by nurses who perform both nursing and administrative work than nurses who perform only nursing work.

4 Discussion

First results will be discussed in the light of recent literature, followed by the limitations of this research and recommendations for further research.

4.1 Interpretation of the results

Both the SUS and the QUIS are considered highly reliable instruments (Bangor et al. 2009, Hoyt et al. 2013). This research confirmed the reliability of both instruments with Cronbach's alpha's of 0.81 (SUS) and 0.98 (QUIS).

The median [interquartile interval] SUS score in this research was 67.50 [57.50 – 85.00]. The tendency for SUS scores between 0 – 100 to be perceived as percentages is incorrect (Brooke 2013). A study of Bangor et al. (2009) shows that with a mean SUS score of 50,9 (sd= 13,8) we can talk about 'OK' usability and with a mean SUS score of 71,4 (sd = 11,6) we can talk about 'good' usability (Bangor et al. 2009). Several studies interpret their SUS scores by this method (Fritz et al. 2012, Sousa et al. 2016, Bakhshi-Raiez et al. 2012), while other researchers report that a SUS score higher than 68 means the software is acceptable by industry standards (Nair et al. 2015, Ahn et al. 2016).

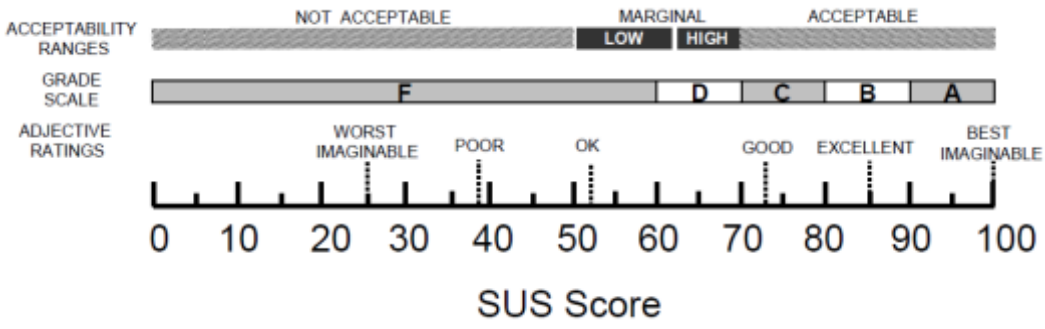

Figure 1. From "Determining what individual SUS scores mean: adding an adjective rating scale" by A. Bangor, P. Kortum, J. Miller, 2009, Journal of usability studies.

The QUIS scores are more difficult to interpret because different scores were obtained on different levels (as a whole, per factor and per item). There is no guideline on what QUIS scores mean but the developers of the QUIS state that the midpoint of the rating scale (5) can be used as a criterion. If the item is

above 5, it can be perceived as being better than an arbitrary, mediocre value (Norman and Schneiderman 2017). The overall median [interquartile interval] in this research is 6,50 [5,50;7,50] and ratings for all usability items are higher than neutral (5), resulting in a positive usability score (Gadd et al. 2011, Staggers et al. 2007).

On the basis of our SUS and QUIS scores we can conclude that the investigated software packages used in Flanders have a usability that ranges from acceptable to good. They are effective, efficient and satisfying. This is a surprisingly positive finding because previous research showed that usability of ENRs is scored relatively low by nurses (Cho et al. 2016, Topaz et al. 2016). Although that previous research showed relatively low usability scores by nurses, better results are found in evaluations by other health care personnel. For example, two studies show positive QUIS results of an EMR and EHR. A first study where students evaluate an EHR had a mean QUIS score of 7.40 (Hoyt et al. 2013). A second study showed QUIS scores of the interface factors all above 7.00 of an EMR assessed by physicians (Jaspers et al. 2008), which demonstrates that although ENRs are scored relatively low by nurses, studies show positive results by other healthcare personnel which with further research can hopefully also be noted in usability research of ENRs. No studies were found specifically evaluating an EMR, EHR, or ENR with the SUS.

The positive results found in our research may also mean that software vendors in Flanders were able to meet usability standards, contrarily to research in the U.S. that shows that some vendors fail to meet usability standards (Ratwani et al. 2017). This assumption is premature, further research evaluating the adherence of software vendors to usability standards will be necessary.

In addition to the individual SUS and QUIS scores, a correlation was sought between SUS and the first QUIS interface factor ‘overall user reactions’ and between the SUS and the median QUIS score. To the best of our knowledge no previous research combined the two questionnaires in evaluating software of EHRs, ENRs and EMRs. Therefore, we wanted to test if there was a correlation

between the SUS, which measures the overall usability of the software, and the first interface actor of the QUIS, which measures the overall user reactions, OR if there was a correlation between the SUS and the overall median of the QUIS. Indeed, positive correlations were found between both. But only a low correlation r_s=0.433 (p=0.001) was found between the SUS and the first interface actor. A significantly higher but still only moderate correlation r_s=0.608 (p=0.000) was found between the SUS and the overall median of the QUIS, meaning that the SUS is better correlated with the overall median of the QUIS than with the first interface actor of the QUIS.

Subsequently, correlations were sought between the SUS and QUIS scores and three sociodemographic variables (age, work experience in years and number of months using software). No correlations were found between the SUS and QUIS interface factors and the three sociodemographic variables. And only a few low but significant correlations were found between QUIS items and the three sociodemographic variables. Both ‘age’ and ‘work experience in years’ were positively correlated with two QUIS items ‘error messages (unhelpful-helpful)’ and ‘system warns you about potential problems (never-always)’. ‘Number of months using the software’ was contrarily to the other sociodemographic variables negatively correlated to four QUIS items, which shows an interesting finding that how longer you use the software, how lower you score the usability of the software. We can suspect however that, although low correlations were found for a few QUIS items, our research shows that age, work experience in years and number of months using software are not of great influence on our software’s usability. In only one other study a Spearman rank test is reported exploring a correlation between time-motion scores and the total QUIS score. In that study tasks had to be completed before filling in the QUIS (Hoyt et al. 2013). A study of Bangor showed a significant correlation between the SUS score and age but it was only a small correlation (r- and p-values were not mentioned). No other studies report correlations between the SUS or QUIS scores and sociodemographic variables. This could be because most of the research combines the SUS or QUIS with a qualitative approach and interprets only the overall SUS and QUIS scores (Fritz et al. 2012, Nair et al. 2015, Ahn et

al. 2016), or SUS and QUIS scores are used to compare different prototypes (Narvaez et al. 2016, Nair et al. 2015, Jaspers et al. 2008), systems (Ahn et al. 2016) or user groups (Sousa et al. 2016).

At last several significant differences were found between the SUS and type of work, between the QUIS factor 'system capabilities' and type of work, and between several QUIS items and type of work. Which means that the type of work influences the usability results, more specifically nurses who perform only nursing work give significantly lower scores on the usability of the software than nurses who perform nursing and administrative work. Only one other study compared their QUIS score with sociodemographic variables but no significant differences were found (Hoyt et al. 2013), making our results showing differences according to type of work a new finding in usability research of ENRs.

4.2 Limitations

Although the results of this research are positive, several limitations have to be taken into account. There was not a lot of research available concerning the usability of ENRs, making this work exploratory. Therefore, interpretations were based on the literature available and on extra literature that was found in other domains where usability research was performed. This led to some limitations in comparing our results with the literature because not a lot of articles were available in this specified context of use.

Several types of bias may have occurred during this research. A first one is sampling bias. Distortion of results may have occurred due to adding a convenience sample to the randomised sampling results. This choice was made due to a low response rate (12%) of the randomised sample. Quantitative usability research is known to be mostly performed on small sample sizes, ranging from 13 to 47 participants in the available studies (Nair et al. 2015, Narvaez et al. 2016, Bakhshi-Raiez et al. 2012, Sousa et al. 2016, Ahn et al. 2016, Staggers et al. 2007). Studies who do have big sample sizes mostly report low response rates varying from 27% to 41% (Jaspers et al. 2008, Hoyt

et al. 2013). Although there was a low response rate this should not have influenced our survey results.

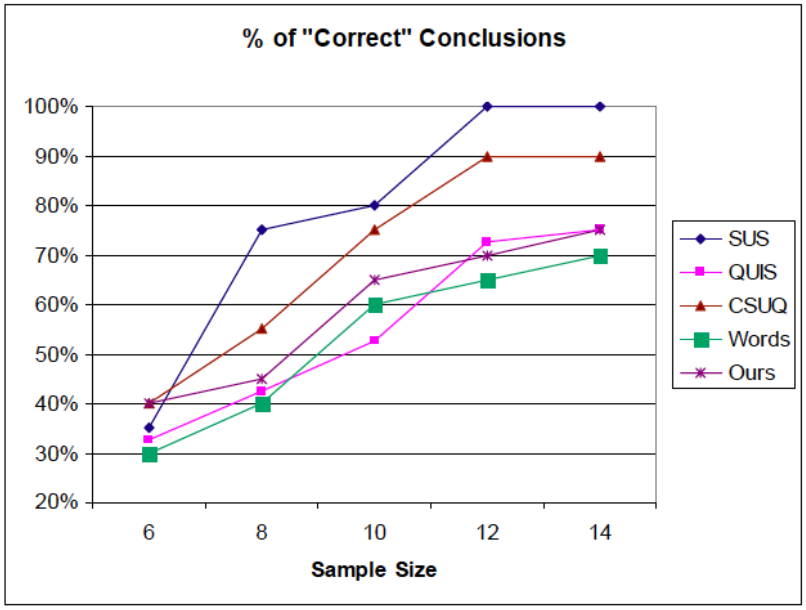

Figure 2. From "A comparison of questionnaires for assessing website usability" by T.S. Tullis and J.N. Stetson, 2004, Proceedings of UPA 2004 Conference.

Figure 2 shows that SUS reaches asymptotes of 90-100% of 'correct conclusions' with a sample size of 12 for assessing website usability. The improvement by going to a sample size of 14 is small. QUIS reaches asymptotes of 70-75%, meaning that the QUIS would have continued to yield improvement if larger samples than 14 had been tested in the research (Tullis and Stetson 2004). Our QUIS sample ranged from 30-45 differing from the questions, which probably led to higher asymptotes than a sample of 14. The developers of the QUIS state that a sample size of at least 20 is necessary for statistical purposes (Norman and Schneiderman 2017).

Most of the studies, including ours, performed usability evaluations on working systems but it would actually be better if the studies applied usability methods within the specification/development phase (Peute et al. 2008, Nair et al. 2015, Horsky et al. 2010). Iterative usability evaluation during the development stages makes the product more specific to users' needs (Yen and Bakken 2012). Researchers should keep in mind that usability research needs to start during

the different stages of product development, which in this research was not possible because the software packages were already in use.

There was chosen not to compare the software packages in this study to protect the anonymity of the software vendors. However, other studies show that the comparison between different systems can shed an interesting light on usability and give more meaning to the SUS and QUIS scores (Narvaez et al. 2016, Ahn et al. 2016, Jaspers et al. 2008).

Several studies only evaluated certain components of the software, which gave more depth to the usability evaluation (Rose et al. 2005, Staggers et al. 2007, Nykanen et al. 2012), while we opted to evaluate the whole system as a starter of usability research in Flanders.

Other studies let the participants perform several tasks before evaluating the system or software by means of a questionnaire (Staggers and Rodney 2012, Hoyt et al. 2013).

In the literature study conducted previous to writing this thesis it was shown that there are a lot of different study designs that can be used to evaluate the usability of EHRs, EMRs or ENRs, ranging from quantitative to qualitative to mixed-method designs. In this study there was opted to conduct a quantitative research, although several other methodologies were possible. No decision can be made on which usability method is better than another based on the literature available. There is a lack of formal and standardized reporting of usability evaluation results.

4.3 Recommendations for further research

The abovementioned limitations lead to the recommendations for further research. It seems best to opt for smaller sample sizes to improve the response rate. It can also be interesting to evaluate different aspects of the registered nursing software, for example: medication, prescriptions, patient information, etc. to gain more specific knowledge on the usability of different aspects of the ENRs. Using a mixed-method or qualitative approach can also bring a new and

interesting influence, comparing it with the evaluation of only a part of the software. Another possibility is to let the participants perform several tasks before evaluating the software. As is shown in this thesis, this is an exploratory study, leading to a lot of different possibilities for further research.

5 Conclusion

Usability is an upcoming domain of interest, resulting in a need for research on this topic. Results of this exploratory study in the field of usability of ENRs in home care nursing are encouraging. Both the SUS and QUIS scores show ok to good usability of the registered ENRs in Flanders, with a median SUS score of 67.50 [57.50 – 85.00] and an overall QUIS score of 6.50 [5.50 - 7.50]. Although low correlations were found for a few QUIS items, we can suspect that our research shows that age, work experience in years and number of months using software are not of great influence on our software's usability. Several significant differences were found between the SUS and QUIS and type of work, showing that nurses who perform only nursing work give lower usability scores than nurses who perform both administrative and nursing work. Further research is necessary to deepen our knowledge on usability of the registered ENRs used by home care nurses in Flanders.

References

(2017) Registratie van de medische softwarepakketten. e-health platform.

Ahn, C., Cho, Y., Oh, J., Song, Y., Lim, T.H., Kang, H. & Lee, J. (2016) Evaluation of Smartphone Applications for Cardiopulmonary Resuscitation Training in South Korea. *Biomed Res Int,* **2016**, 6418710.

Bakhshi-Raiez, F., de Keizer, N.F., Cornet, R., Dorrepaal, M., Dongelmans, D. & Jaspers, M.W. (2012) A usability evaluation of a SNOMED CT based compositional interface terminology for intensive care. *Int J Med Inform,* **81**(5), 351-62.

Bangor, A., Kortum, P. & Miller, J. (2009) Determining what individual SUS scores mean: adding an adjective rating scale. *J. Usability Studies,* **4**(3), 114-123.

Brooke, J. (1996) SUS - A quick and dirty usability scale. *Usability evaluation in industry,* **189**(194), 4-7.

Brooke, J. (2013) SUS: a retrospective. *J. Usability Studies,* **8**(2), 29-40.

Charles, D., Gabriel, M. & Searcy, T. (2015) Adoption of electronic health record systems among U.S. non-federal acute care hospitals: 2008-2014. *ONC Data Brief,* **23**.

Cho, I., Kim, E., Choi, W.H. & Staggers, N. (2016) Comparing usability testing outcomes and functions of six electronic nursing record systems. *Int J Med Inform,* **88**, 78-85.

Ellsworth, M.A., Dziadzko, M., O'Horo, J.C., Farrell, A.M., Zhang, J. & Herasevich, V. (2017) An appraisal of published usability evaluations of electronic health records via systematic review. *J Am Med Inform Assoc,* **24**(1), 218-226.

Finstad, K. (2006) The system usability scale and non-native English speakers. *J. Usability Studies,* **1**(4), 185-188.

Fritz, F., Balhorn, S., Riek, M., Breil, B. & Dugas, M. (2012) Qualitative and quantitative evaluation of EHR-integrated mobile patient questionnaires regarding usability and cost-efficiency. *Int J Med Inform,* **81**(5), 303-13.

Gadd, C.S., Ho, Y.X., Cala, C.M., Blakemore, D., Chen, Q., Frisse, M.E. & Johnson, K.B. (2011) User perspectives on the usability of a regional health information exchange. *J Am Med Inform Assoc,* **18**(5), 711-6.

Garets, D. & Davis, M. (2017) EMRs and EHRs. EuroRec: European institute for health records.

Horsky, J., McColgan, K., Pang, J.E., Melnikas, A.J., Linder, J.A., Schnipper, J.L. & Middleton, B. (2010) Complementary methods of system usability evaluation: Surveys and observations during software design and development cycles. *Journal of Biomedical Informatics,* **43**(5), 782-790.

Hoyt, R., Adler, K., Ziesemer, B. & Palombo, G. (2013) Evaluating the Usability of a Free Electronic Health Record for Training. *Perspectives in Health Information Management / AHIMA, American Health Information Management Association,* **10**(Spring), 1b.

Jansen-Kosterink, S.M., Wever, D., Hermens, H.J. & Vollenbroek-Hutten, M.M.R. (2012) Differences in use of a exercise-based tele-rehabilitation service delivered as substitute of or supplement to conventional care.

Jaspers, M.W., Peute, L.W., Lauteslager, A. & Bakker, P.J. (2008) Pre-post evaluation of physicians' satisfaction with a redesigned electronic medical record system. *Stud Health Technol Inform,* **136**, 303-8.
Lewis, J.R. & Sauro, J. (2009) The Factor Structure of the System Usability Scale. In *Proceedings of the 1st International Conference on Human Centered Design: Held as Part of HCI International 2009*Springer-Verlag, San Diego, CA, pp. 94-103.
Nair, K.M., Malaeekeh, R., Schabort, I., Taenzer, P., Radhakrishnan, A. & Guenter, D. (2015) A Clinical Decision Support System for Chronic Pain Management in Primary Care: Usability testing and its relevance. *J Innov Health Inform,* **22**(3), 329-32.
Narvaez, S., Tobar, A.M., Lopez, D.M. & Blobel, B. (2016) Human-Centered Design of an mHealth App for the Prevention of Burnout Syndrome. *Stud Health Technol Inform,* **228**, 215-9.
Norman, K.L. & Schneiderman, B. (2017) QUIS: Questionnaire for User Interaction Satisfaction. Vol. 2017 University of Maryland.
Nykanen, P., Kaipio, J. & Kuusisto, A. (2012) Evaluation of the national nursing model and four nursing documentation systems in Finland--lessons learned and directions for the future. *Int J Med Inform,* **81**(8), 507-20.
Page, C.A. & Schadler, A. (2014) A nursing focus on EMR usability enhancing documentation of patient outcomes. *Nurs Clin North Am,* **49**(1), 81-90.
Peute, L.W., Spithoven, R., Bakker, P.J. & Jaspers, M.W. (2008) Usability studies on interactive health information systems; where do we stand? *Stud Health Technol Inform,* **136**, 327-32.
Ratwani, R.M., Fairbanks, R.J., Hettinger, A.Z. & Benda, N.C. (2015) Electronic health record usability: analysis of the user-centered design processes of eleven electronic health record vendors. *J Am Med Inform Assoc,* **22**(6), 1179-82.
Ratwani, R.M., Hettinger, A.Z. & Fairbanks, R.J. (2017) Barriers to comparing the usability of electronic health records. *J Am Med Inform Assoc,* **24**(e1), e191-e193.
Riskin, L., Koppel, R. & Riskin, D. (2015) Re-examining health IT policy: what will it take to derive value from our investment? *J Am Med Inform Assoc,* **22**(2), 459-64.
Rose, A.F., Schnipper, J.L., Park, E.R., Poon, E.G., Li, Q. & Middleton, B. (2005) Using qualitative studies to improve the usability of an EMR. *J Biomed Inform,* **38**(1), 51-60.
Scandurra, I. & Liljequist, D. (2016) Ten Demands of Improved Usability in eHealth and Some Progress - Co-Creation by Health and Social Care Professionals. *Stud Health Technol Inform,* **225**, 227-31.
Schoen, C., Osborn, R., Squires, D., Doty, M., Rasmussen, P., Pierson, R. & Applebaum, S. (2012) A survey of primary care doctors in ten countries shows progress in use of health information technology, less in other areas. *Health Aff (Millwood),* **31**(12), 2805-16.
Sockolow, P.S., Liao, C., Chittams, J.L. & Bowles, K.H. (2012) Evaluating the Impact of Electronic Health Records on Nurse Clinical Process at Two Community Health Sites. *NI 2012: Proceedings of the 11th International Congress on Nursing Informatics,* **2012**, 381.

Sousa, V.E., Lopes, M.V., Ferreira, G.L., Diniz, C.M., Froes, N.B. & Sobreira, B.A. (2016) The construction and evaluation of new educational software for nursing diagnoses: a randomized controlled trial. *Nurse Educ Today,* **36**, 221-9.

Staggers, N., Kobus, D. & Brown, C. (2007) Nurses' evaluations of a novel design for an electronic medication administration record. *Comput Inform Nurs,* **25**(2), 67-75.

Staggers, N. & Rodney, M. (2012) Promoting usability in organizations with a new health usability model: implications for nursing informatics. *NI 2012 (2012),* **2012**, 396.

Sturm, J., Bakx, I., Cranen, B., Terken, J. & Wang, F. (2002) Usability evaluation of a Dutch multimodal system for train timetable information. In *LREC 2002*Gran Canaria de Las Palmas pp. 255 - 261.

Topaz, M., Ronquillo, C., Peltonen, L.M., Pruinelli, L., Sarmiento, R.F., Badger, M.K., Ali, S., Lewis, A., Georgsson, M., Jeon, E., Tayaben, J.L., Kuo, C.H., Islam, T., Sommer, J., Jung, H., Eler, G.J., Alhuwail, D. & Lee, Y.L. (2016) Nurse Informaticians Report Low Satisfaction and Multi-level Concerns with Electronic Health Records: Results from an International Survey. *AMIA Annu Symp Proc,* **2016**, 2016-2025.

Tullis, T.S. & Stetson, J.N. (2004) A comparison of questionnaires for assessing website usability. In *Proceedings of the Usability Professionals Association (UPA) 2004 Conference*, pp. 7-11.

Yen, P.Y. & Bakken, S. (2012) Review of health information technology usability study methodologies. *J Am Med Inform Assoc,* **19**(3), 413-22.

Zahabi, M., Kaber, D.B. & Swangnetr, M. (2015) Usability and Safety in Electronic Medical Records Interface Design: A Review of Recent Literature and Guideline Formulation. *Hum Factors,* **57**(5), 805-34.

Appendix 1

System Usability Scale

1. Ik denk dat ik deze software graag regelmatig willen gebruiken.
2. Ik vind de software onnodig complex.
3. Ik vind de software makkelijk te gebruiken.
4. Ik denk dat ik ondersteuning nodig heb van een technisch persoon om deze software te kunnen gebruiken.
5. Ik vind dat de verschillende functies van deze software erg goed geïntegreerd zijn.
6. Ik vind dat er teveel tegenstrijdigheden in de software zitten.
7. Ik kan me voorstellen dat de meeste mensen zeer snel leren om deze software te gebruiken.
8. Ik vind deze software erg omslachtig in gebruik.
9. Ik voelde me erg vertrouwd met de software.
10. Ik moest erg veel leren voordat ik met deze software aan de slag kon.

Questionnaire for User Interaction Satisfaction

Overall user reactions

1.1	Algemene reacties op de software:	verschrikkelijk schitterend 1 2 3 4 5 6 7 8 9	n.v.t.
1.2		frustrerend bevredigend 1 2 3 4 5 6 7 8 9	n.v.t.
1.3		saai stimulerend 1 2 3 4 5 6 7 8 9	n.v.t.
1.4		moeilijk makkelijk 1 2 3 4 5 6 7 8 9	n.v.t.
1.5		inadequate power adequate power 1 2 3 4 5 6 7 8 9	n.v.t.
1.6		rigide flexibel 1 2 3 4 5 6 7 8 9	n.v.t.

Screen

2.1	Scherm layouts waren nuttig	nooit altijd 1 2 3 4 5 6 7 8 9	n.v.t.
2.1.1	Hoeveelheid informatie die weergegeven kan worden op het scherm	ongeschikt geschikt 1 2 3 4 5 6 7 8 9	n.v.t.
2.1.2	Ordening van informatie op het scherm	onlogisch logisch 1 2 3 4 5 6 7 8 9	n.v.t.
2.2	Opeenvolging van schermen	verwarrend duidelijk 1 2 3 4 5 6 7 8 9	n.v.t.
2.2.1	Volgende scherm in een opeenvolging	onvoorspelbaar voorspelbaar 1 2 3 4 5 6 7 8 9	n.v.t.
2.2.2	Teruggaan naar het vorige scherm	onmogelijk makkelijk 1 2 3 4 5 6 7 8 9	n.v.t.
2.2.3	Opeenvolging van werk gerelateerde taken	verwarrend duidelijk aangeduid 1 2 3 4 5 6 7 8 9	n.v.t.

Terminology and system information

3.1	Gebruik van terminologie in het systeem	inconsequent 1 2 3 4 5 6 7 8 9 consequent	n.v.t.
3.1.1	Werkgerelateerde terminologie	inconsequent 1 2 3 4 5 6 7 8 9 consequent	n.v.t.
3.1.3	Computer terminologie	inconsequent 1 2 3 4 5 6 7 8 9 consequent	n.v.t.
3.2	Berichten die op het scherm verschijnen	inconsequent 1 2 3 4 5 6 7 8 9 consequent	n.v.t.
3.2.1	Positie van instructies op het scherm	inconsequent 1 2 3 4 5 6 7 8 9 consequent	n.v.t.
3.3	Berichten die op het scherm verschijnen	verwarrend 1 2 3 4 5 6 7 8 9 duidelijk	n.v.t.
3.3.1	Instructies voor commando's en functies	verwarrend 1 2 3 4 5 6 7 8 9 duidelijk	n.v.t.
3.3.2	Instructies voor het corrigeren van fouten	verwarrend 1 2 3 4 5 6 7 8 9 duidelijk	n.v.t.
3.4	Computer houdt u op de hoogte over wat hij doet	nooit 1 2 3 4 5 6 7 8 9 altijd	n.v.t.
3.4.1	Geanimeerde cursors houden u op de hoogte	nooit 1 2 3 4 5 6 7 8 9 altijd	n.v.t.
3.4.2	Het uitvoeren van een taak leidt tot een voorspelbaar resultaat	nooit 1 2 3 4 5 6 7 8 9 altijd	n.v.t.
3.4.4	Feedback mogelijkheden instellen	onmogelijk 1 2 3 4 5 6 7 8 9 makkelijk	n.v.t.
3.4.5	Duur van vertraging tussen operaties	onaanvaardbaar 1 2 3 4 5 6 7 8 9 aanvaardbaar	n.v.t.
3.5	Foutmeldingen	onnuttig 1 2 3 4 5 6 7 8 9 nuttig	n.v.t.

Learning

4.1	Leren om de software te bedienen	moeilijk makkelijk 1 2 3 4 5 6 7 8 9	n.v.t.
4.1.1	Beginnen	moeilijk makkelijk 1 2 3 4 5 6 7 8 9	n.v.t.
4.1.2	Leren van geavanceerde functies	moeilijk makkelijk 1 2 3 4 5 6 7 8 9	n.v.t.
4.1.3	Tijd om de software te leren gebruiken	traag snel 1 2 3 4 5 6 7 8 9	n.v.t.
4.2	Taken kunnen uitgevoerd worden op een duidelijke manier	nooit altijd 1 2 3 4 5 6 7 8 9	n.v.t.
4.2.1	Aantal stappen per taak	Te veel Juist genoeg 1 2 3 4 5 6 7 8 9	n.v.t.
4.2.2	Stappen om een taak te voltooien volgen een logische volgorde	nooit altijd 1 2 3 4 5 6 7 8 9	n.v.t.
4.2.3	Feedback over de voltooiing van stappen	onduidelijk duidelijk 1 2 3 4 5 6 7 8 9	n.v.t.

System capabilities

5.1	Software snelheid	te traag	snel genoeg	
		1 2 3 4 5 6 7 8 9		n.v.t.
5.1.1	Responstijd voor de meeste taken	te traag	snel genoeg	
		1 2 3 4 5 6 7 8 9		n.v.t.
5.1.2	Informatie over de snelheid wordt weergegeven	te traag	snel genoeg	
		1 2 3 4 5 6 7 8 9		n.v.t.
5.2	De software is betrouwbaar	nooit	altijd	
		1 2 3 4 5 6 7 8 9		n.v.t.
5.2.1	Operaties zijn	onbetrouwbaar	betrouwbaar	
		1 2 3 4 5 6 7 8 9		n.v.t.
5.2.2	Optreden systeemstoringen	frequent	zelden	
		1 2 3 4 5 6 7 8 9		n.v.t.
5.2.3	De software waarschuwt u voor potentiële problemen	nooit	altijd	
		1 2 3 4 5 6 7 8 9		n.v.t.
5.3	Verbeteren van uw fouten	moeilijk	makkelijk	
		1 2 3 4 5 6 7 8 9		n.v.t.
5.3.1	Verbeteren typfouten	complex	eenvoudig	
		1 2 3 4 5 6 7 8 9		n.v.t.
5.3.2	Mogelijkheid om operaties ongedaan te maken	ongeschikt	geschikt	
		1 2 3 4 5 6 7 8 9		n.v.t.
5.4	Bedieningsgemak hangt af van uw niveau van ervaring	nooit	altijd	
		1 2 3 4 5 6 7 8 9		n.v.t.
5.4.1	Je kan taken volbrengen door slechts enkele commando's te kennen	met moeite	gemakkelijk	
		1 2 3 4 5 6 7 8 9		n.v.t.
5.4.2	Je kan functies/snelkoppelingen gebruiken	met moeite	gemakkelijk	
		1 2 3 4 5 6 7 8 9		n.v.t.

Sociodemographic questionnaire

1. Geslacht

o Man

o Vrouw

2. Leeftijd:

3. Van welke softwarepakket maakt u gebruik?

Welke versie?

Hoelang maakt u er reeds gebruik van (in maanden)?

4. Maakte u voor dit softwarepakket reeds gebruik van een ander software pakket? Ja/Nee

Zoja, welk?

5. Werkt u zelfstandig of in een organisatie (vb. Wit-Gele kruis,…)?

6. Welk werk verricht u?

o Uitsluitend administratief werk

o Uitsluitend verpleegkundig werk

o Beiden

7. Wat is uw hoogst behaalde diploma?

o HBO-5 (gegradueerde verpleegkundige)

o Bachelor in de verpleegkunde

o Bachelor na bachelor

o Postgraduaat

o Master

o Master na master

Specifieer:

8. Hoelang bent u reeds werkzaam (in het algemeen als verpleegkundige)?