

**USABILITY TESTING OF THE WEB ACCESSIBLE POPULATION
PHARMACOKINETICS HEMOPHILIA SERVICE (WAPPS-HEMO)**

**THE USABILITY TESTING OF THE WEB ACCESSIBLE POPULATION
PHARMACOKINETICS HEMOPHILIA SERVICE (WAPPS-HEMO)**

by

ISLAM BARGHASH, BPHARM

A Thesis Submitted to the School of Graduate Studies in Partial Fulfillment of the
Requirements for the Degree Master of Science in eHealth

McMaster University © Copyright by Islam Barghash, September 2015

McMaster University MASTER OF SCIENCE IN EHEALTH (2015) Hamilton, Ontario

TITLE: The Usability Testing of the Web Accessible Population Pharmacokinetics
Hemophilia Service (WAPPS-Hemo)

AUTHOR: Islam Barghash, BPharm

SUPERVISOR: Dr. Alfonso Iorio

NUMBER OF PAGES: xiii, 96

Supervisory Committee

The Usability Testing of the Web Accessible Population Pharmacokinetics Hemophilia
Service (WAPPS-Hemo)

By

Islam Barghash, BPharm

Supervisory Committee

Dr. Alfonso Iorio, MD, PhD, FRCP(C) – **Supervisor**
Associate Professor, Department of Clinical Epidemiology & Biostatistics
Deputy Chief, Health Information Research Unit (HIRU)

Dr. Ann McKibbin, MLS, PhD FMLA – **Committee Member**
Professor Emeritus, Department of Clinical Epidemiology & Biostatistics

Dr. Anne Holbrook, MD, PharmD, MSc, FRCP(C) – **Committee Member**
Professor, Division of Clinical Pharmacology and Toxicology, Department of Family
Medicine
Director, Division of Clinical Pharmacology and Toxicology

Abstract

Introduction

Hemophilia is a genetic disorder that is caused by deficiencies in coagulation factor VIII and factor IX. Optimal management of hemophilia requires tailoring the dose of treatment to the individual patient's need. This tailoring is based on several clinical considerations, for example, bleeding phenotype and desired level of activity, and estimated individual pharmacokinetic (PK) parameters. While a classical PK approach would require several post infusion blood samples taken over multiple days, a population PK approach might enable individual assessment using fewer samples. Health information technologies can support implementation of sophisticated, easily available, point-of care resources to estimate PK values with a population approach. The Web Accessible Population Pharmacokinetic Service-Hemophilia (WAPPS-Hemo) is a proposal for such a solution, developed at McMaster University. Once tested, it will be offered to hemophilia centres worldwide.

The objective of the study was to investigate the ease of use (usability) of the WAPPS-Hemo web interface among clinicians and other people who treat hemophilia through two iterative cycles of usability testing.

Methods

The total number of participants was 13, and they were physicians, nurses and research coordinators. The think aloud technique was selected for testing to gain feedback and comments on the participants' thought processes while interacting with the system

and discover interface design problems. Additionally, the System Usability Scale (SUS) questionnaire was used to obtain data on user satisfaction.

Results

The initial assessment of the prototypal WAPPS-Hemo interface with SUS reported a score of 70.5, which is considered an above average score. We received many useful suggestions through two iterations of user testing, ending with a final SUS score of 73 after implementation of the suggested improvements. Verbal feedback from users in the second round showed that users experienced an easier and more intuitive interaction with the system.

Conclusion

Usability testing and analyses were conducted in this study to discover user interface issues and to determine the usability and learnability of the WAPPS-Hemo service among various potential users. Through iterative cycles, application of the think aloud technique, and the SUS questionnaire, we optimized the usability of the WAPPS-Hemo program and have moved to implementation (June 2015).

Dedication

To my dear father, mother, and my beloved wife.

Acknowledgments

I would like to thank the following:

Dr. Alfonso Iorio, for his insights, guidance and experience.

Dr. Ann Mckibbon, for her inspiration throughout the entire eHealth program and the thesis writing.

Dr. Anne Holbrook, for her support and push for improvements.

Dr. Arun Keepanasseril, for help and support.

Tamara Navarro, for her help in recruitments and facilitations of the usability sessions.

The IT team for their contribution in enhancing the WAPPS-Hemo.

Iris Kehler, for her support and guidance throughout the eHealth program.

Table of Contents

Supervisory Committee	iii
Abstract	iv
Dedication	vi
Acknowledgments	vii
Table of Contents	viii
List of Tables	xi
List of Abbreviations.....	xiii
CHAPTER 1: INTRODUCTION	1
CHAPTER 2: LITERATURE REVIEW	4
2.1 Hemophilia	4
2.2 Population Pharmacokinetics in Hemophilia	6
2.3 Usability of Health Information Systems	7
2.4 User-Centred Design of Health Information Systems	9
2.5 Usability Testing Methods	11
2.5.1 The Think Aloud Method	13
2.5.2 System Usability Scale (SUS)	17
2.5.3 Remote Usability Testing	20
CHAPTER 3: METHODS	23
3.1 The WAPPS-Hemo	23
3.2 Study Design	24
3.3 Selection of Participants	24
3.4 Usability Testing Procedure	26
3.5 Background Questionnaire	27
3.6 Usability Testing Session	27
3.7 Testing Procedure and Plan	27
3.7.1 Study Introduction and Description	27
3.7.2 The Think Aloud Technique	28
3.7.3 SUS Questionnaire	29
3.7.4 Post-Test Debriefing	29
3.7.5 Data Collection	30
3.7.6 Data Analysis	30

3.7.7	First Iteration	31
3.7.8	Second Iteration	31
4.1	First Iteration	32
4.1.1	Pre-test Background and Demographics: First Iteration	32
4.1.2	Usability	33
4.1.3	Interface Problems	35
4.1.4	SUS	45
4.1.5	Age	47
4.1.6	Session Elapsed Times	47
4.2	User Interface Modifications	48
4.3	Second Iteration	56
4.3.1	Pre-test Background and Demographics: Second Iteration	56
4.3.2	Usability	58
4.3.3	Interface Problems	60
4.3.4	SUS	67
4.3.5	Age	69
4.3.6	Session Elapsed Times	69
4.3.7	Experienced Users	70
4.4	Summary of Results	72
CHAPTER 5: DISCUSSION AND FUTURE WORK		73
5.1	Discussion	73
5.2	Limitations	77
5.3	Implications and Future Work	79
CHAPTER 6: CONCLUSION		82
REFERENCES		83
Appendix A: Task List		89
Appendix B: Background Questionnaire		90
Appendix C: Introduction Script		91
Appendix D: SUS Survey		93
Appendix E: Post-Testing Debriefing		94
Appendix F: Summary of Results		95

Appendix G: NVivo 10 Coding for Iteration 1..... 96
Appendix H: NVivo 10 Coding for Iteration 2..... 96

List of Tables

		Page
Table 1	System Usability Scale (SUS) Items.....	18
Table 2	Understanding the System Usability Scale (SUS) Score.....	19
Table 3	Iteration 1- Demographic Descriptors of Participants (n=5)..	32
Table 4	Iteration 1- Usability Issues.....	34
Table 5	Iteration 1- Suggestions for Improvement of User Interface	44
Table 6	Iteration 1- System Usability Scale Mean Scores.....	46
Table 7	Iteration 1- Usability Sub-Scale Scores.....	46
Table 8	Iteration 1- Individual SUS Scores based on Age.....	47
Table 9	Iteration 1- Usability Session Times (minutes).....	48
Table 10	Iteration 1- Suggestions for Improvement and Implementations	49
Table 11	Iteration 2- Demographic Descriptors of Participants (n=8)..	57
Table 12	Iteration 2- Usability Issues.....	59
Table 13	Iteration 2- Suggestions for Improvements of User Interface	66
Table 14	Iteration 2-System Usability Scale Mean Scores.....	68
Table 15	Iteration 2- SUS Sub-scale Scores.....	68
Table 16	Iteration 2- Individual SUS Scores based on Age.....	69
Table 17	Iteration 2- Usability Session Times (in minutes).....	70
Table 18	Iteration 2- SUS Scores of Experienced Users	71

Table 19	Iteration 2- SUS Sub-scale Scores Experienced Users.....	71
Table 20	Iteration 2- Individual SUS Scores based on Age (Experienced Users).....	72
Table 21	Iteration 2- Usability Session Times (in minutes) of Experienced Users.....	72
Table 22	Summary of SUS Scores.....	72

List of Abbreviations

BMI:	Body mass index
HIRU:	Health Information Research Unit
ISO:	International Organization for Standardization
ISTH:	International Society on Thrombosis and Haemostasis
IT:	Information technology
PK:	Pharmacokinetic
SDLC :	Software Development Life Cycles
SUS:	System Usability Scale
WAPPS-Hemo:	Web Accessible Population Pharmacokinetic Service – Hemophilia

CHAPTER 1: INTRODUCTION

Hemophilia is a rare genetic disease caused by the deficiency of one of the factors of the clotting cascade. Clotting factors are proteins required for normal blood clotting. These proteins work with platelets to help in the clotting process. Platelets are small blood cells that help blood clot. When blood vessels are damaged, the clotting factors help platelets stick together and form plugs to stop the bleeding. There are two types of hemophilia, A and B. Types A and B are characterized by the deficiency of factors VIII and IX, respectively. Excessive bleeds and hemorrhages occur due to the inactivity of the coagulants. The hemorrhages usually cause internal bleeding in elbows and knees that puts pressure on the joints resulting in severe pain. Additionally, hemorrhages can cause internal bleeding in muscles that cause limbs to swell and result in pain (Gringeri, Lundin, Von Mackensen, Mantovani, & Mannucci, 2011; Valentino et al., 2012; “What Is Hemophilia?,” 2013). Prophylaxis, the prevention of bleeding, is accomplished by the infusion of clotting factor concentrates to people with hemophilia. The infusion of clotting factor concentrates are given at specific doses and at a frequency such that plasma level of the deficient proteins are maintained above the threshold needed to prevent bleeding. Studies show an increase in life expectancy of patients with hemophilia because they undergo prophylaxis with these infusions (Wang et al., 2015). Factor VIII or IX concentrates come from human plasma or genetically engineered using recombinant DNA technology (“Factor Replacement Therapy,” 2015).

The ideal dosing of the factor concentrates is heavily reliant on patients’ individual pharmacokinetic (PK) characteristics. PK dose tailoring may significantly

contribute the management of hemophilia. However, standard calculation of PK parameters for an individual requires several sample blood draws to achieve accurate and appropriate dosing. For example, patients with hemophilia B require seven blood sample draws over a 72-hour period. (Collins, Fischer, Morfini, Blanchette, & Björkman, 2011). To facilitate dosing estimates, data from a population of patients with hemophilia could be collected and used in the application of Bayesian-based estimates to assess individual PK parameters starting from a population PK model.

Today, health information systems in clinical settings play an integral role in the management of diseases and patient care. However, user problems and usability issues often arise (in one study, 64% of the time), making these systems demanding to operate. Difficult to use systems can become obstacles to care efficiencies rather than enablers of clinical improvements (Teixeira, Ferreira, & Santos, 2012).

Usability is a characteristic that measures the ease of use of system interfaces and includes methods to improve and enhance user experience during the development process (Nielsen, 2012). Usability testing can be employed in any phase of the lifetime of a system but is most often done during development. According to the International Organization for Standardization's (ISO) definition of usability, it is the degree in which specific users can use a product to attain specific goals with effectiveness, efficiency and satisfaction (P.-Y. Yen & Bakken, 2012).

The following chapters proceed as follows: a literature review on the relevant topics pertaining to the study, the methodologies and usability testing procedures,

elucidation of the results, and finally the discussion and interpretation of the results and recommendations for future work.

CHAPTER 2: LITERATURE REVIEW

2.1 Hemophilia

Hemophilia A and hemophilia B are rare, X-chromosome-linked recessive genetic diseases caused by the deficiency of factor VIII and factor IX in the blood. For people with hemophilia their coagulant activity is not sufficient and causes hemorrhages (Valentino et al., 2012). Recurrent hemarthroses is the most common hemorrhagic manifestation of hemophilia, which are mainly in the elbows, knees and other joints. This joint involvement can lead to progressive destruction of the joints, irreversible arthropathy and chronic pain (Gringeri et al., 2011). Recurrent bleeding into joints is a major cause of morbidity (Valentino et al., 2012). Therefore, the main goal for these patients is to prevent arthropathy by replacing the deficient coagulation factor with concentrates (Gringeri et al., 2011).

Empirical therapy with prophylactic schedules with factor VIII was developed and tested in clinical trials. The basic idea, proposed initially by the Swedish treaters, was a consequence of observing that only severe patients (those with circulating factor level $<0.01-0.02$ U/ml) bleed, while mild and moderate (those with levels above $0.01-0.02$ U/ml) almost never have spontaneous bleeding. Hence, prophylaxis was conceived as periodical administration of factor concentrates at a dose and a frequency to ensure the deficient factor concentration was kept above the $0.01-0.02$ U/ml threshold (Nilsson, Hedner, & Ahlberg, 1976). A simple clinical observation, and some PK calculation have been at the base of prophylactic treatment hypothesis. Clinicians started to recommend

that approach, which soon became the standard of care, even though data from controlled trials were not conclusively established until very recently (Manco-Johnson et al., 2007). In a study of children with haemophilia by Gringeri, prophylaxis was defined as the continuous administration of concentrates in a long-term, regular fashion, based on two or three infusions on a weekly basis, although sometimes more than three infusions per week were required. Based on this yet limited evidence, clinicians unanimously consider the prophylactic administration of factor concentrates the most effective methods to predict bleeding and its complications, and prophylaxis is recommended by authoritative guidelines (Srivastava et al., 2013). Studies found that prophylaxis with recombinant factor VIII prevented hemarthroses and structural joint damage effectively in patients with hemophilia A (Aledort, Haschmeyer, & Pettersson, 1994; Soucie et al., 2001). That approach reduced the risk of developing joint damage and showed improved quality of life of young patients, especially in children, with severe factor VIII deficiency (Manco-Johnson et al., 2007; Valentino et al., 2012). However, inadequate prophylaxis may occur due to the difficulty in obtaining venous access in children, insufficient patient and family compliance with the proposed treatments. (Gringeri et al., 2011) Additionally, the large inter-individual variability of the dose response to factor concentrate infusion, which makes it difficult to meet the goal of keeping the circulating factor level above the 0.01-0.02 UI/ml efficacy threshold. This is where PK have its role in optimizing prophylaxis in hemophilia.

2.2 Population Pharmacokinetics in Hemophilia

Indeed, the plasma concentration of a coagulation factor is largely related to its hemostatic efficacy. The dose, the time of infusion, and the patient's PK response to the dosing all help determine the plasma concentration of the coagulation level. PK measurement is time-consuming and tiring in clinical practice. According to the International Society on Thrombosis and Haemostasis (ISTH) guidelines, patients with hemophilia require the following four sampling periods: at the time of pre-infusion (baseline), 10-15 minutes after the completion of the infusion, and 30 minutes and one hour later. Patients with hemophilia A would require additional sampling at 3, 6, 9, 24, 28, and 32 hours post-infusion; and possibly a 48-hour sample if the patient was given 50 IU/kg (Lee M, Morfini M, Schulman S, 2001). That lengthy process puts more stress and pressure on patients and parents of young patients that can negatively affect treatment adherence. According to Sheiner et al, the Bayesian estimation method can reduce the number of blood samples required (Sheiner et al, 1979) per patient when estimating the PK properties of a drug. Specifically, estimating PK properties from sparse data samples in a large population was initially developed to assess PK variability in population difficult to study, such as in pediatric oncology. The basic goal in PK estimation is to model the variability in the population by using as much information as possible from easy to collect covariates, such as age, weight, body mass index (BMI), and laboratory values.

Initially a population PK model is been built, either starting from a large population of patients with sparse data, or from a sufficiently large sample of patients

undergoing a classical PK study. With these data, a Bayesian post-estimation method can be used as a statistical basis to estimate the PK of an individual patient based on minimum blood sample draws and information from the underlying population.

For haemophilia, Björkman stated that three key parameters be taken into consideration to estimate PK parameters for individual patients using a Bayesian approach: body weight, dosing specification, and a minimum of two post-infusion plasma drug concentration measurements (Björkman, 2010). Additional variables for each patient with hemophilia A are blood group, von Willebrand factor levels, and hematocrit. For patients with hemophilia B an additional variable is creatinine clearance.

This three-parameter approach can assist calculating the half-life of the coagulation factor from data gathered from two or three blood draws. Hence, tailoring PK dosing is essential in clinical practice where the convenience and reduced number of sampling is required for clinicians and patients. (Collins et al., 2011)

2.3 Usability of Health Information Systems

Many health information systems have sophisticated design and have been developed for clinical settings. However, their usefulness and ease of use can be questionable. It is thought that health information systems fail because of the lack of the involvement of potential end users during the development process and because of non-technology related problems (Britto et al., 2009; Teixeira, Ferreira, & Santos, 2007).

Additionally, poor usability can cause frustration among clinicians, errors, and increased risk of compromising patient safety (Ratwani, Fairbanks, Hettinger, & Benda, 2015).

Several definitions of system usability have been described. Nielsen mentioned that usability does not have a single, one-dimensional property of a system user interface (Nielsen, 1993). The ISO defines usability as the degree in which specific users can use a product to attain specific goals with effectiveness, efficiency, and satisfaction (Po-Yin Yen, Sousa, & Bakken, 2014). Moreover, a system is considered usable when a user is able to accomplish tasks in a safe, efficient, effective and enjoyable manner (A. W. Kushniruk & Patel, 2004; Peute, Spithoven, Bakker, & Jaspers, 2008). Nielsen further clarifies that by defining five elements of usability: learnability, efficiency, memorability, errors and satisfaction (Nielsen, 1993).

Appropriate user or human centred designs and engaging users during the development process can enhance the efficiency, ease of use, learning, and understandability of a health information system. Productivity, adoption rate, and user satisfaction will likely be enhanced when potential users are involved during the development phases. In addition, medical errors and the cost of training new users might be reduced (A. W. Kushniruk & Patel, 2004; Teixeira et al., 2007). The benefits of allowing users and stakeholders to participate in the design process are to obtain useful feedback from users and provide the opportunity to resolve any problems, such as website interface issues, as early as possible before the final product is finalized (J. Stinson et al., 2014; Teixeira et al., 2012).

Without proper user engagement, a health information system can be perceived as cumbersome and difficult to learn and use, which leads to high resistance among users, particularly with physicians and nurses (Teixeira et al., 2007). User frustration can increase and the probability of users leaving the entire system will likely increase as well (J. Stinson et al., 2014; Teixeira et al., 2007). According to Teixeira et al, lack of user involvement, unrealistic expectations, and vague goals are some of the main reasons why health information systems fail They are responsible for the failure of health information systems 13%, 6% and 5% of the time, respectively (Teixeira et al., 2012; Teixeira, Saavedra, Ferreira, & Sousa Santos, 2009).

User engagement permits better understanding of human factors and interactions involved in using a particular system. Ultimately, the proper use of a health information system can have the potential to improve patient safety and the quality of health care (Khajouei & JasperS, 2010; Teixeira et al., 2007). To assure patient safety, health information systems must undergo extensive usability tests and support user-centred health care processes (Scandurra, Hägglund, Persson, & Ahlfeldt, 2014).

2.4 User-Centred Design of Health Information Systems

In systems design of eHealth projects, such as in health information systems, constant evaluation in the development process is necessary to ensure safety, effectiveness, and efficiency (M. W M Jaspers, 2009; A. Kushniruk, 2002). In a typical software system development occurs in a set of fixed stages ending with a final stage of

system evaluation. Iterative design can form the basis of software development life cycles (SDLC).

One of the most common problems with health information technology (IT) adoption is usability factors. Health IT initiatives provide support to multiple functions and a range of multidisciplinary users giving rise to challenges in designing an ideal system (P.-Y. Yen & Bakken, 2012). An example of an internet intervention where usability was considered is the Teens Taking Charge: Managing Arthritis Online treatment program. Its designers suggested that it was essential to include usability testing during its development. In addition, feedback was collected from potential users to understand what worked, what did not work, and missing features or functions through iterative cycles to improve the prototype (Jennifer Stinson et al., 2010). As suggested by Kushniruk, evaluation pushes forward continual development and design of a system, and the most powerful of evaluation methods is usability testing (Kushniruk, 2002). Clinicians not providing enough attention to designing a health IT might cause a system to become inefficient, ineffective, and unsatisfactory. As explained by Wolpin, development has been mainly based on what health care providers, researchers, and vendors thought was appropriate for the end-user, without focusing on user-centred design (Wolpin et al., 2014). However, lack of input and evaluation including usability testing is remediable if developers focus on health IT usability and engage the target users, such as physicians, nurses or pharmacists (P.-Y. Yen & Bakken, 2012).

Usability can be evaluated through a series of iterations in SDLC. Iteration comprises of the repetition of system design and development phases to improve and

fine-tune the system in each round. The testing process starts during the initial development of the system and then goes through evaluation, gaining feedback from users to enhance and polish the system. Input from users often leads to further evaluation cycles (A. Kushniruk, 2002).

That cyclical evaluation approach can improve understanding, increase the frequency of use, and help determine how likely potential end users will implement the system (Jennifer Stinson et al., 2010). It is necessary for applications and web-based systems, during their design and development, to include constant and iterative feedback and system adaptation from potential users.

2.5 Usability Testing Methods

Usability testing of information systems with end users to determine how they interact and analyze the technology can provide insights into how useful that technology might be (Daniels, Fels, Kushniruk, Lim, & Ansermino, 2007; A. Kushniruk, 2002). For example, in the Teens Taking Charge: Managing Arthritis Online treatment web application, the user performance, ease of use, learnability, errors and efficiency, user satisfaction, user interface design, and functionality were measured (Jennifer Stinson et al., 2010). Usability should be tested and evaluated during a product's development cycle to have the potential to meet needs end and provide what users expect, especially for improvement of user interface design (Kushniruk, 2002; Yen & Bakken, 2012).

There have been a number of methods established to evaluate the usability of systems. Initially, some approaches relied on summative evaluation to assess a system

after its development. The goal of summative evaluation is to meet predefined criteria to discover functionality problems, safety, efficiency, and the system's impact on outcome measures, such as cost. Formative evaluation was introduced to assess usability during system development. The goal of formative evaluation is to identify usability issues in an iterative approach throughout the development lifecycle of a system. (A. W. Kushniruk & Patel, 2004) System usability testing methods can be expert-based or user-based. Expert-based inspection method involves the use of heuristic evaluation and cognitive walkthroughs, while the user-based testing method involves user performance measurements, cognitive workload assessments, satisfactory questionnaires, and interviews. (M. W M Jaspers, 2009)

Heuristic evaluation involves assessing a system's user interface against a set of recognized usability principles. Monkman et al. have applied heuristic evaluation testing to a mobile blood pressure tracking application (Monkman, Griffith, & Kushniruk, 2015). The evaluation included eight essential heuristics for health literacy and usability and three optional heuristics for specific content or device. On the other hand, the cognitive walkthrough method is used to evaluate a system's learnability and a user's interaction with the system's interface (M. W M Jaspers, 2009).

The user-based testing method incorporates the think aloud method. The think aloud method requires subjects to speak out loud in relation to what they are doing while performing tasks or trying to solve issues that may arise during their usage of a system (M. W M Jaspers, 2009; Virzi, 1992). The method is also referred to as direct method (Hsieh & Shannon, 2005). The user-based testing method also can include the assessment

of subjective finding from users through a series of questionnaires or survey. For example, the System Usability Scale (SUS) can be used to gather subjective information from users after interacting with a system (Brooke, 1996).

2.5.1 The Think Aloud Method

Think aloud data provide information on the ongoing thought process by tapping in to a participant's cognitive processes. For example, in the study by Jaspers et al, the oncologists were given a set of instructions to follow and tasks to complete. They were encouraged to constantly speak out loud while interacting with the system. Although at a minimum, testers reminded the oncologist to continue to speak out loud and verbalize their interactions if they become silent (Monique W M Jaspers, Steen, van den Bos, & Geenen, 2004). Participants' comments were recorded, transcribed, and analyzed qualitatively. Participants were encouraged to speak about their observations, experiences, and interactions with the system while performing assigned tasks. The target interventions could include, but were no limited to, free exploration of a web site or program and completing a short list of tasks to carry during the usability testing. This approach can help uncover errors in navigation and provide other observations or insights (Wolpin et al., 2014).

There are four types of think aloud techniques: the a) Concurrent and b) Retrospective Think Aloud, and the c) Concurrent Probing and d) Retrospective Probing ("Running a Usability Test," 2014). The Concurrent Think Aloud technique is used to understand a person's thought process while they are thinking out loud during their

interaction with a system. The advantages of this approach are that the tester can understand the participant's thoughts while interacting with the system, revealing any problems that may arise. The person being observed can also provide spontaneous feedback and emotional responses. The Retrospective Think Aloud method simply involves asking participants to review and repeat their steps after the usability session and its tasks are complete. With this technique, the usability metrics, such as time elapsed and user satisfaction are not affected by the need to talk aloud. However, the retrospective technique may increase session duration and provide inaccurate data. The Concurrent Probing method allows the tester to ask questions while the participant provides comments and feedback during the tasks. The advantage to this approach is that it allows the tester to comprehend the participants' thoughts as they interact with the system and are working through the tasks. With Concurrent Probing the natural thought process of participants is interrupted and their progression through the tasks might be hindered. Finally, the Retrospective Probing requires testers to wait until the completion of the usability session and they then start to ask questions and follow up on participants' comments, concerns, and questions the person had doing the tasks. This technique does not interfere with the usability metrics (e.g., the time taken for each task). However, data collection might not be accurate or adequate. ("Running a Usability Test," 2014)

The think aloud technique has been previously applied on a hemophilia self-management website (Breakey et al., 2013) and on pediatric oncology systems (Monique W M Jaspers et al., 2004). Breakey et al's approach involved the tester observing participants using an interactive haemophilia website designed to assist patients

transitioning from pediatric to adult care. Iterative cycles of the usability testing were conducted by the study observer to determine the usability and perceptiveness of the website's user interface. The think aloud approach provided data for analysis and obtained a clearer picture of the cognitive processes as problems occurred while the user used the website.

Before starting actual testing, pediatric oncologists were given instructions about the think aloud technique and what it involved. They used paper-based samples of patient records to practice the verbalization of their thoughts (Monique W M Jaspers et al., 2004). In other studies, such as the one conducted by Cristancho-Lacroix, participants were not guided but rather given written instructions to follow with respect to talking aloud. Unlike the study by Breakey, a camera was used in the oncology study to record the order in which the pediatric oncologists looked for patient information. Participants in the Cristancho-Lacroix study were also videotaped for behavioural analysis.

These think aloud methods were originally focused on user interface problems in IT studies. However, the same methods can be used to assess user satisfaction and the effectiveness of implementation of clinical guidelines. Hence, usability testing can be ideal to gain hands on experience and feedback from clinicians on the applicability of a system for medical practice (A. Kushniruk, 2002).

Furthermore, the think aloud approach was also used in the usability testing and development of a web based cancer symptom and quality-of-life support intervention (Wolpin et al., 2014). Individual sessions used an eye tracker to track users' eye movements and fixations on certain areas of the website in addition to the usability

methods described above. This approach can yield accurate data and useful feedback, but is often not done because of system requirements for capturing and analyzing eye tracking data.

Task durations and total time spent on tasks and sessions vary from one person to the other and with the complexity of the tasks and what is expected from participants. During the hemophilia self-management study (Breakey et al., 2013), the participants were guided to the main menu of the hemophilia management website and were asked to select a module that interested them. Then the participants were directed to certain key features within the module. Even though the sessions had a time limit of 45-60 minutes, sessions lasted from 20 minutes to 90 minutes.

The think aloud technique was also used for caregivers of patients with Alzheimer's disease (Cristancho-Lacroix et al., 2014). The web application tested was designed to decrease stress of the caregivers, who are often family members of these patients. The objective of the usability study was to describe the development process of a web-based program for informal caregivers of patients with Alzheimer's disease and to improve the user-centered design.

To assess user performance, the performance of participants was observed as they navigated through the website taking advantage of its contents and features. Errors and navigation obstacles were recorded. User satisfaction was assessed through a series of questions that comprised of "like" or "dislike" responses (Breakey et al., 2013). In a similar context, a five point Likert scale was used in the study by Cristancho-Lacroix, which was 0 = negative (dislike) to 4 = positive (like). The aim of this survey was to assess

five topics: 1) the users' overall impression about the website, 2) ease of use, 3) how pleasant it was to use the website, 4) how coherent the layout was, and 5) how satisfied the users were about the design of the website.

The study by Jaspers used two questionnaires, the QUIIS-short version (Questionnaire for User Interaction Satisfaction), which is a standardized questionnaire for user interface satisfaction, and a self-developed questionnaire (Monique W M Jaspers et al., 2004).

2.5.2 System Usability Scale (SUS)

A number of surveys can help designers uncover usability problems with their systems or products. The SUS can help evaluate systems and different interface types (Bangor, Kortum, & Miller, 2009). The SUS (Arsand & Demiris, 2008) is a set of questions that produces reliable results across different samples. The questionnaire measures the overall user satisfaction. However, it can be more specific in terms of “ease of use, ease of learning, simplicity, effectiveness, information and user interface” (Johnson, Johnson, & Zhang, 2005). The SUS scale is scored on a 5-point Likert scale, where 5 = strongly agree and 1 = strongly disagree (Brooke, 1996). One of the features that makes the SUS appealing is that it consists of ten questions only (**Table 1**). Participants can complete it in a short time. The final score can be calculated quickly and ranges from 0 to 100, making it easy to understand by the respective stakeholders (Bangor et al., 2009).

The SUS was developed by Digital Equipment Corporation© in 1986. Three general usability criteria are defined by the ISO 9241-11 (Borisci, Federici, & Lauriola, 2009). The criteria are measuring effectiveness, in which potential users are able to use the system and perform the intended tasks and the quality of the output of these tasks; measuring efficiency, in which potential users are able to perform the required tasks with least resources; and measuring user satisfaction, which comprises of potential users' experiences, reactions, comfort, and feedbacks while using the system and completing the tasks (Borisci et al., 2009; Brooke, 2013).

It is vital for a web based system to be easy to use, fit the workflow of the prospective users' daily tasks, and offer useful information. Usability is critical in designing clinical informatics applications (Wakefield, Pham, & Scherubel, 2015) as health care needs useful, effective, and accurate systems to support patient care. Information provided by the system, icons, and instructions must be easy use and provided in a user-friendly language (Wakefield et al., 2015).

Table 1: System Usability Scale (SUS) Items
1. I think that I would like to use this system frequently.
2. I found the system unnecessarily complex.
3. I thought the system was easy to use.
4. I think that I would need the support of a technical person to be able to use this system.
5. I found the various functions in this system were well integrated.
6. I thought there was too much inconsistency in this system.

7. I would imagine that most people would learn to use this system very quickly.
8. I found the system very cumbersome to use.
9. I felt very confident using the system.
10. I needed to learn a lot of things before I could get going with this system.

2.5.2.1 SUS Score and Calculation

The SUS produces a single number that represents the measure of the overall usability of the system being tested. However, the scores of the individual items of the SUS are not meaningful on their own. Readers who are interested in the method for calculating SUS scores are referred to the papers by Bangor and Brooke (Bangor et al., 2009; Brooke, 1996). In **Table 2**, the average SUS scores correspond to their respective explanations.

Table 2: Understanding the System Usability Scale (SUS) Score	
Adjective	Mean SUS Score
Worst Imaginable	12.5
Awful	20.3
Poor	35.7
OK	50.9
Good	71.4
Excellent	85.5
Best Imaginable	90.9

Adapted from (Bangor et al., 2009; Kobak et al., 2011)

2.5.3 Remote Usability Testing

There are advantages and disadvantages of remote usability testing over the traditional method of usability testing. According to a study by Thompson, remote testing may be at least as effective as traditional testing (Thompson, Rozanski, & Haake, 2004). One benefit is to avoid high costs of travel of test participants, especially those who are not close to the testing facility. Additionally, remote testing could be useful to the tester to know and see how the test system performs on other computer systems (Schade, 2013). For example, hospitals and hemophilia centres might have computers with different configurations, restrictions, and specific networks. Additionally, these systems might have older versions of internet browsers installed (Internet Explorer, Google Chrome, or Firefox).

Many approaches have been used to perform qualitative usability testing with semi-structured interviews based on watching individuals interact with a system while completing tasks related to that system (Peute et al., 2008). The interviews, when possible are audio and video recorded. A researcher can be present to observe and take notes (Jennifer Stinson et al., 2010).

It is essential to design and redesign a system to suit potential end-users. The information needed for these design cycles includes recording and analysing users,

identifying usability problems such interface frustration and navigation inconsistency, the time taken to perform a task on the web system, and suggestions for improvement (Kushniruk, 2002).

We think that the think aloud method is the most appropriate tool to assess the usability of the WAPPS-Hemo. Since this method is a user-based method, the results of the usability testing through random potential end users might provide a better, unbiased scope of the system's usability than compared to the expert method. Therefore, the results of this method could potentially address issues that are likely to be faced by end users. Participants will have the opportunity to speak their mind while interacting with the WAPPS-Hemo, which gives the tester a chance to understand potential user problems without limiting participants to predefined set of possible usability issues (heuristic evaluation). Additionally, we thought the SUS could be used to quantitatively measure potential end user satisfaction. The SUS survey is simple to use, quick to complete, and provides overall user satisfaction score of the WAPPS-Hemo. This could allow the tester to quantitatively compare the usability of the system from one iteration to another.

The objective of this study was to enhance the user interface design of the WAPPS-Hemo service through a series of two usability testing iterations. These iterations have been conducted to simulate a clinical setting to represent a real-world situation in a hemophilia clinic. We wanted to discover user interface issues that participants encountered while interacting with the system. The WAPPS-Hemo research team

anticipated that clinician and other user acceptance and ease of use would improve after testing the second version of the WAPPS-Hemo service.

The research questions proposed for this study are as follows:

1. To identify user interface design issues that health care providers who work with patients who have haemophilia encountered when they use the WAPPS-Hemo service.
2. To collect user suggestions on improvements to the WAPPS-Hemo system.
3. To determine the level of user satisfaction health care providers report after using the WAPPS-Hemo service.

CHAPTER 3: METHODS

In this section, the web app service, the study design and procedure, participant selection, data collection, and analysis will be described.

3.1 The WAPPS-Hemo

The WAPPS-Hemo is an existing web service that allows clinicians, other care health care providers, and administrators, to perform individual PK assessments for the treatment of patients with hemophilia in an easy and simplified manner. The web service is a centralized and dedicated web service that generates certified hemophilia patient PK data, where the data are actively moderated and manually validated. The assessment will focus on the treatment of adults and children with hemophilia. The service provides individual PK estimation with a smaller number of blood samples compared to the conventional method. Another useful benefit of the service is that upon switching from one coagulation factor to another, it will simplify the PK parameter calculations for the new coagulation factor. These automated population PK estimations are intended to provide optimal therapeutic treatments and prophylaxis to people with hemophilia. (Iorio, 2014)

Population PK and their Bayesian statistical assessments are the underlying engines of a PK assessment system being developed at McMaster University, the (WAPPS-Hemo). The service allows clinicians, other health care providers, research

coordinators, and administrators to perform individual PK assessments for the treatment of patients with hemophilia A and B in an easy and simplified manner.

3.2 Study Design

The study was conducted on an early prototype of the WAPPS-Hemo service. The first part of the study was to collect qualitative data on user interaction with the service, user impressions and feedback done in a laboratory setting. This data collection is comprised of comments, opinions, assistance requested, and overall design issues subjects encountered during the usability testing sessions. The second part comprised of completing the SUS questionnaire to collect subjects' satisfaction scores of the system. In addition, the time to complete each session of the usability testing was also recorded. Section 3.7 provides more information on the actual steps taken in conducting the usability testing.

3.3 Selection of Participants

Small sample sizes can be used in usability analysis and testing. Nielson suggested that 80% of usability problems can be identified with four or five subjects (Nielsen & Landauer, 1993), hence four to six participants were tested in the study by Breakey and eight in the study by Jaspers. Usability testing can produce informative

results with small sample sizes. Cristancho-Lacroix included six caregivers (clinicians and families) participating in their usability testing.

Participants of the study are clinicians from hemophilia-treating centres who are registered and participated in the WAPPS-Hemo service and associated non-clinicians. To insure we had participants with relevant backgrounds and roles within a hemophilia-treating centre we included nurses, clinical research coordinators, data managers, and administrators. We do not know who will be using and inputting patient data from the hemophilia-treating centres into the system being tested. The choice of participants varied in each of the two iterations. We also recruited participants from cities across Ontario and other provinces (Alberta and British Columbia), and the United States. Depending on the availability and accessibility, the preferred group of participants were of different backgrounds. We recruited potential participants using email invitations. We felt it was valuable to obtain qualitative and quantitative data and feedback on the WAPPS-Hemo from different perspectives and viewpoints. Participants were given identification numbers to ensure anonymity.

The Hamilton Integrated Research Ethics Board approved the WAPPS-Hemo project (14-601-D).

The usability testing sessions took place in the Health Information Research Unit (HIRU) at McMaster University. Since most of the initial participants were off-site, the testing sessions were conducted remotely. However, if online testing was not possible due to technical difficulties, participants were switched to on-site testing. For the remote

testing, we used Join.Me to connect with the subjects. Join.Me allows audio and screen recording. For the on-site testing we used ActivePresenter software to capture audio and screen interactions.

3.4 Usability Testing Procedure

The approach taken to conduct the usability testing was built on the user-based testing method (M. W M Jaspers, 2009), which included iterative cycles and questionnaires, similar to a study by Korus (Korus, Cruchley, Stinson, Gold, & Anthony, 2015).

The usability testing was conducted on an early prototype of the WAPPS-Hemo application, where the interface was tested rather than the PK engine functionality and performance characteristics (precision, accuracy, reproducibility) outcome. The testing was performed over two iterative cycles, the first iteration was carried out on the initial version of the WAPPS-Hemo service. The second iteration was done after the suggested changes were applied (revised version). In each iteration, subjects were given fabricated hemophilia data for two patients to test the system with three tasks (Appendix A). These tasks simulated a real-world situation of patients' blood infusions, blood sample draws, and plasma factor levels.

3.5 Background Questionnaire

Prior to each session, subjects were asked to fill in a background questionnaire. Participants were asked several questions pertaining the nature of their role at their respective hemophilia centres, if they had any experience with the WAPPS-Hemo service, and if they used or were currently using a PK calculator. This survey was done online on SurveyMonkey. (Appendix B)

3.6 Usability Testing Session

Subjects were registered onto the WAPPS-Hemo system prior starting the usability testing. The testing was conducted on a test server that was identical to the production server. The data entered were anonymous patient data.

3.7 Testing Procedure and Plan

In the following sections, I describe the process taken to conduct the usability sessions.

3.7.1 Study Introduction and Description

After the subject filled the background questionnaire online, the author explained the usability testing and described what was expected from the subject during the session. A script adapted and modified according to the time allocated for each participant and

the facilities available, was used (Rubin & Chisnell, 2008), Appendix C. This explanation comprised of explaining the think aloud technique and encouraging subjects to speak out loud throughout their interaction with the WAPPS-Hemo. Subsequently, the explanation of the usability test was followed by a semi-structured interview to produce further feedback about their impressions and potential improvements they may suggest for the future of the WAPPS-Hemo.

3.7.2 The Think Aloud Technique

The subjects' interactions with the WAPPS-Hemo started with the login screen. From there they had to start entering patient's PK data. The usability testing sessions were carried out by the author, but the WAPPS-Hemo project coordinator was present. In addition, the author took field notes during the sessions to capture any observations of note. For example, the sections of the WAPPS-Hemo in which participants needed author intervention and assistance were noted.

Since talking out loud while using an application can be unnatural to most people, subjects were reminded during the sessions with prompts adopted from Wolpin's study. For example, "Tell me what you are thinking" and "Keep talking" (Wolpin et al., 2014). Almost every subject was given a different set of patient information. This approach was important to obtain greater insight on how subjects can interpret various data types and apply them. The subjects were asked open ended questions at the end of the usability testing. They also usually provided some suggestions for improvements as they navigated

throughout the system. These suggestions were expected because withholding opinions and first impressions was not natural to most subjects.

3.7.3 SUS Questionnaire

After subjects had completed the assigned tasks carrying out the think aloud technique, they were provided with a link to the SUS questionnaire (Appendix D) online on SurveyMonkey's website. The author asked the subjects to immediately fill the SUS questionnaire to acquire the best possible and realistic answers on their satisfaction with the WAPPS-Hemo service.

3.7.4 Post-Test Debriefing

After the session, the subjects were asked to give their opinion and overall feelings on the WAPPS-Hemo service and its features (Appendix E). This review was intended for the subjects to wrap up their thoughts about the system and point out any final obstacles or issues. Usually, valuable spontaneous feedback was given at this point. This semi-structured interview was conducted to discover other aspects of user satisfaction and recommendations for improving the website.

3.7.5 Data Collection

The following data were collected during the study:

The sessions were held privately in a testing room where the author and the project's coordinator were present. First, basic background information was collected to help identify the subjects' professions and demographic information. Second, subjects' comments, opinions and observations were audio and screen recorded, in addition to written field notes. Third, open ended responses were collected for further descriptive analysis. Finally, the subjects' SUS responses were gathered. The patient profiles used as test tasks contained data that were similar to what will be entered and reviewed through a typical patient encounter in a hemophilia-treating centre. Participants were asked to read the tasks and then refer to the patient profiles while being encouraged to speak out loud as they progress with each task.

3.7.6 Data Analysis

The directed content analysis (Hsieh & Shannon, 2005) approach was taken for the qualitative method.

The usability session recordings were imported to NVivo 10 software. The qualitative sessions were transcribed verbatim and were time-stamped. Following the work of Graham (Graham et al., 2008), coding categories have been developed. These categories represented different themes to correspond to the various types of usability

problems. In addition, simple content analysis was performed on the results of each iterative cycle, as explained by Sandelowski (Sandelowski, 2000). The results and feedback from the iterative sessions helped in identifying themes that were similar and consistent across subjects (Korus et al., 2015; Nolan et al., 2015).

3.7.7 First Iteration

Qualitative analysis was conducted to identify usability problems of the system. A directed content analysis approach was taken to organize that data (Graham et al., 2008; Hsieh & Shannon, 2005). This approach helped in predicting a preliminary coding scheme. The codes developed are interface problems, technical problems, suggestions for improvements, positive, negative and neutral comments. The texts from the transcriptions were referenced to these categories according to their appropriate descriptions. (Appendix F) The data then were exported to Microsoft Excel and consolidated. In Microsoft Excel, the comments from the sample were color coded to identify frequency of similar comments and user situations.

3.7.8 Second Iteration

The approach to conducting and analyzing data from the second iteration were similar to the first iteration.

CHAPTER 4: RESULTS

4.1 First Iteration

This first iteration took approximately four weeks to complete.

4.1.1 Pre-test Background and Demographics: First Iteration

The background questionnaire that was conducted prior commencing the testing indicated that of the five participants, two participants were physicians and/or hematologists (40%), two were registered nurses or nurse practitioners (40%) and one was an administrator (20%). In addition, only one participant (20%) was using a PK calculator at the time of the testing and none of the participants had prior experience with the WAPPS-Hemo service. (Table 3)

Table 3: Iteration 1- Demographic Descriptors of Participants (n=5)			
		n	%
Gender	Male	2	40
	Female	3	60
Age (years)	Mean	40.2	
Role in Treating Centre	Physician/Hematologist	2	40
	RN/Nurse Practitioner	2	40
	Administrator	1	20
Experience	Use of any PK calculator	1	20
	WAPPS-Hemo	0	0

4.1.2 Usability

Five subjects, who are potential users of the service, participated in the first iteration of the usability testing. Two subjects were men and three were women with a mean age of 40.2 years. A list of user interface problems is presented in **Table 4**. This list identifies the interface design items that subjects had trouble with and the number of occurrences. With respect to redesign of the system, the higher the occurrences were, the higher the priority was given to attempt changes applied by the programmers. Interface problems were defined as user interface and design issues and misconceptions. Technical problems are defined as issues that appear to users as back-end (server-side) restrictions.

Suggestions for improvements were coded. (**Appendix G**) During the process of thinking out loud, subjects provided their opinions and visions on how certain aspects or functionality of the WAPPS-Hemo service could be enhanced. These comments were identified and listed under the suggestion for improvements category. In addition, subjects had the opportunity to suggest any other improvements they desire after they had finished the required tasks.

Table 4: Iteration 1- Usability Issues		
Interface Problems		
ID	Item	Number of Occurrences
1	Confusion between the Finalize Measurements button and Add Measurement button under Collected Data section	5
2	The Add Infusion button under Recommended Fields confuses subjects into adding a second infusion for the same patient	4
3	Drop-down (expand) button was not clear to subjects that they had to go there to reveal the blood sample draws entry fields	4
4	The Drug drop-down menu was perceived to contain all the drugs, but subjects did not know where to navigate to add and activate specific drugs	3
5	There is no option to select positive or negative blood group under the Blood Group drop-down	2
6	There is no indication of what unit to use in the Infusion Duration field	2
7	After adding an infusion, the populated fields become blank, confuses subject into enter the data again, unnecessarily	2
8	Warning signs are not visible enough to be noticed	1
Technical Problems		
1	After adding an infusion, subjects are frustrated they cannot go back to edit a mistake in the infusion profile	4
2	Subject was frustrated that Infusion Duration would not accept numbers with decimals (e.g. 5.5)	1

4.1.3 Interface Problems

Interface problems were referred to as obstacles and difficulties a user encounters with the interface design. User interface design focuses on what potential users might need to perform certain tasks. Therefore, it is critical to ensure that users can understand and perform these tasks easily (“User Interface Design Basics,” 2014).

All of the participants in the sample had a common issue with the sample blood draws data entry section. As seen in **Figure 1**, the Finalize Measurements button and the Add Measurement button’s placement were confusing to all participants. At that stage in the web application, subjects did not know what to do or what the next step was. One subject mentioned that they wanted to save or store the infusion information they just entered. However, it was not clear to him or her where a *save* button was. Even though messages appeared when the cursor was hovered above a button, this feature was not intuitive to subjects. A subject referred to the greyed area that provided the ID of the person who had the infusion, ID #242 as seen in Figure 1, commented: “...nowhere that says edit...”. In addition, the subject continued: “...why cannot I highlight?” referred to the same section as mentioned above. Another participant thought that the Add

Figure 1 Blood Sample Data Entry Section

Collected Data

+/-	ID	Drug	BW (kg)	Total U	U/kg	Infusion end time	Duration (mins)	Test	Standard	Opt.	Notes
▼	242	Humate P	80	4000	50.0	2015-04-17 08:00 AM	0	One Stage Coag. (PTT Based)	Drug Specific	☰	
		Measurement Date/Time		Time Elapsed (h:m)		Concentration		Other		Notes	
		2015-04-17 08:00 AM		0:00		1.18		Pre-dose		Edit	
		2015-04-17 08:30 AM		0:30		1.16				Edit	
		<input type="text" value="2015-04-17"/> <input type="text" value="08:00 AM"/> <input type="button" value="← NOW →"/>		TBD		<input type="text"/>		<input type="checkbox"/> Pre-dose <input type="checkbox"/> Below Threshold		<input type="text"/>	
						Finalize Measurements			Add Measurement		
▶	241	Humate P	80	4000	50.0	2015-04-17 08:00 AM	0	One Stage Coag. (PTT Based)	Drug Specific	☰	

Request PK Calculation



Measurement button meant to add another set of infusion data of a new patient. Due to his or her frustration with this situation, a third subject took the initiative and created a new patient profile thinking that there was an error in the system. At that stage, the author had to assist the participants in navigating through the interface to proceed to the next logical step of the task.

The second most common problem that occurred with four participants was the Add Infusion button under the Recommended Fields section as seen below in **Figure 2**. Similar to the aforementioned interface problem, the Add Infusion button seemed to be

out of place and hovering on it did not provide users any immediate hints or assistance. Upon clicking on that infusion button, the patient information entered by the user disappeared, which increased participant frustration. Other participants thought that if they clicked the Add Infusion button, a new infusion of another patient or second infusion data for the same patient would be added. Another participant scrolled straight to the bottom of the page and presumed that clicking the Request PK Calculation button was the next reasonable step to do. However, they received an error by doing this.

Figure 2 Add Infusion Data

Home
Data Entry
Results
Documents
Contact Us
FAQ
Log Out

Patient Data Entry Form
HiRU
[Return to patient selection](#)

Local Patient ID: CM Age: 6 Gender: Male Blood Group: N/A WAPPS ID: 10112

Required Fields

Drug: (v)

Body Weight: (kg)

Total Units Administered: (IU)

Units per kg:

End of Infusion Date: (i)

Time: (i) ⏪ ⏩

Infusion Duration: (i)

Lab Test: (v)

Standard: (v)

Recommended Fields

Height: (cm)

Hematocrit: (%)

Hemoglobin:

Serum Creatinine: (µmol/L)

vWF:RiCof: (µmol/L)

vWF:Ag: (µmol/L)

Notes:

Add Infusion

Collected Data

+/-	ID	Drug	BW (kg)	Total U	U/kg	Infusion end time	Duration (mins)	Test	Standard	Opt.	Notes
▶	229	Kogenate	21	1250	59.5	2015-03-02 08:00 AM	0	One Stage Coag. (PTT Based)	Drug Specific	⊞	

Request PK Calculation

A third interface problem, which came up with the same frequency as the second problem, occurred for four out of five subjects. The Collected Data section is where a user inputs the information from different blood sample draws, as shown in **Figure 3**. However, this section did not expand nor became sufficiently visible for the participant to realize and notice that they had to go to that section. The drop-down arrow was meant for expand and/or collapsing the section, however, the drop-down arrow turned out to cause more confusion than help to first time users of the WAPPS-Hemo service. It was apparent that participants wanted real-time guidance on the system itself to help them navigate and perform the tasks efficiently. For example, to quote a participant regarding that issue, “...totally didn’t pick that up” and “...it could’ve been easier if it would flow all together, I expected it to flow with the rest of the fields”. It was evident that the drop-down icon was not noticeable nor intuitive to most participants.

Figure 3 Drop-down Data Entry

Collected Data										
ID	Drug	BW (kg)	Total U	U/kg	Infusion end time	Duration (mins)	Test	Standard	Opt.	Notes
242	Humate P	80	4000	50.0	2015-04-17 08:00 AM	0	One Stage Coag. (PTT Based)	Drug Specific		
Measurement Date/Time		Time Elapsed (h:m)		Concentration		Other		Notes		
2015-04-17 08:00 AM		0:00		1.18		Pre-dose		Edit		
2015-04-17 08:30 AM		0:30		1.16				Edit		
<input type="text" value="2015-04-17"/> <input type="text" value="08:00 AM"/> <input type="button" value="↕"/>		TBD		<input type="text"/>		<input type="checkbox"/> Pre-dose <input type="checkbox"/> Below Threshold		<input type="text"/>		
<input type="button" value="Finalize Measurements"/>							<input type="button" value="Add Measurement"/>			
▶ 241	Humate P	80	4000	50.0	2015-04-17 08:00 AM	0	One Stage Coag. (PTT Based)	Drug Specific		
<input type="button" value="Request PK Calculation"/>										

Hemophilia-treating centres use different medications. As seen in **Figure 4**, in the Drug field, the drop-down lists various anti-hemophilic factors for a user to choose. Several times participants did not notice the Drug field and continued to populate the text fields, until to a point where they realized that the default chosen drug is not what they used at their respective centres. Participants began to look for a way to edit and change the information. However, the system did not have that edit and change feature. Not surprisingly, this caused annoyances to the participants. One participant commented, “it wasn’t obvious, no one told me I had to activate”. Hence, they required guidance from the author. They continued: “...[there is] no logical design”.

Figure 4 Adding a Drug

Required Fields

Drug:	Advate
Body Weight:	
Total Units Administered:	
Units per kg:	
End of Infusion Date ⓘ:	

- Advate
- Alprolix
- Benefix
- Humate P
- Kogenate**
- Refacto
- Wilate


The former interface problems were the ones with the highest number of occurrences during the usability testing. The following are interface problems that arose with less frequency among these participants.


Figure 5 Blood Group Selection

New Patient Entry

Required Fields

Gender:


Date of Birth :

Local Patient ID :

Consent: Informed consent to enter their data into the system has been received from the patient

Optional Fields

Blood Group:

Baseline Factor Level : (U/ml)

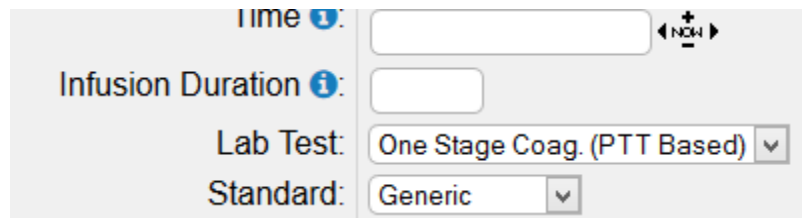
Positive History of Inhibitors:

Upon creating a new patient profile, subjects were presented with drop-down list for the Blood Group, as shown in **Figure 5**. However, a subject wanted to select O+, for example, but was not able to because that blood type was not an option in the drop-down list.

Another minor difficulty encountered was no indication for which unit to use on the Infusion Duration field, as shown in **Figure 6**. If the subject hovered on the “i” icon, which refers to “more information”, they would see the indication, as shown in **Figure 7**. However, from the overall observation and communication with participants, it was

apparent that most people wanted a hint or an indication to be directly visible on the page with the least effort possible.

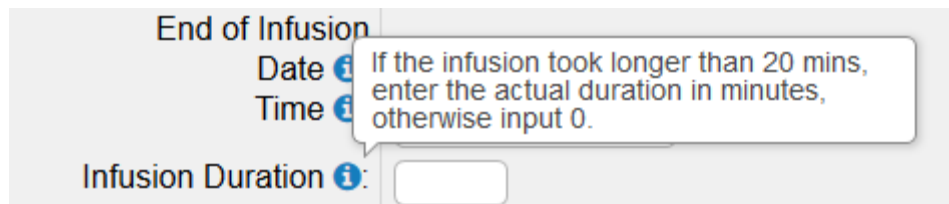
Figure 6 Infusion Duration



The screenshot shows a form section with the following elements:

- A label 'Time' with an information icon (i) and an empty input field with a 'NOW' button.
- A label 'Infusion Duration' with an information icon (i) and an empty input field.
- A label 'Lab Test:' with a dropdown menu showing 'One Stage Coag. (PTT Based)'.
- A label 'Standard:' with a dropdown menu showing 'Generic'.

Figure 7 Infusion Duration Units



The screenshot shows a form section with the following elements:

- A label 'End of Infusion'.
- A label 'Date' with an information icon (i) and an empty input field.
- A label 'Time' with an information icon (i) and an empty input field.
- A tooltip box containing the text: 'If the infusion took longer than 20 mins, enter the actual duration in minutes, otherwise input 0.'
- A label 'Infusion Duration' with an information icon (i) and an empty input field.

Finally, if a participant overlooked a field or did not complete the required information, a warning symbol appeared next to the respective field or item. Only one participant had trouble detecting these warning symbols, although, it became noticeable very quickly with little assistance from the author.

Table 5 shows a consolidated list of improvements suggested after the completion of the first iteration of the usability testing. The suggestions were mainly from the usability interviews, however, they might be slightly modified or improved by the programmers according to the feasibility and appropriateness of the needed change.

Note: ID number corresponds to the Issue item in Table 4.

Table 5: Iteration 1- Suggestions for Improvement of the User Interface	
ID	Item
1	<ul style="list-style-type: none"> • Remove the Add Measurement button • Add a Save button to the right of an active data point field (should disappear after the subject saves the respective entry) • Move the Finalize Measurements button to the centre, above the Request PK Calculation button • A pop-up message when a subject finalizes a measurement, including an OK and Cancel buttons • Request PK Calculation button could be greyed out
2	<ul style="list-style-type: none"> • Replace the Add Infusion with a Save button • A pop-up message when subjects save their entries, including an OK and Cancel buttons • Add a title “Infusion Information” or “List of Available Infusions”
3	<ul style="list-style-type: none"> • When subjects save their entries, the data point fields appear expanded and directly visible • Include a Collapse and Expand button
4	<ul style="list-style-type: none"> • Add a button indicating “Select” or “Choose” button with a link to the centre settings. (with small guide) • If not an admin, pop-up contacting Admin (name, email)
5	<ul style="list-style-type: none"> • Comment Rh not necessary
6	<ul style="list-style-type: none"> • Provide the units in parentheses (e.g. Infusion Duration (min))
7	<ul style="list-style-type: none"> • A pop-up appears indicating that the data have been saved • A message under the Save button that the data have been saved
8	<ul style="list-style-type: none"> • Make warning signs bigger and/or add red warning text

4.1.4 SUS

The SUS item scores were analyzed. As shown in **Table 6**, the mean scores of the participants were calculated for each SUS item/question. To determine the satisfaction level and scores of end-users, the SUS scores were calculated (Brooke, 1996; Wakefield et al., 2015). According to Brooke, a score 68 out of a 100 is considered average. After the data collected were analyzed, the SUS calculated for the first iteration gave a score of 70.5, which is considered above average and according to Bangor et al the score is “Good”. The average scores and standard deviations (SD) were calculated for the usability and learnability items as shown in **Table 7**.

Table 6: Iteration 1- System Usability Scale Mean Scores	
	Subjects (n=5) Mean (SD)^a
1. I think that I would like to use this system frequently.	3 (0.71)
2. I found the system unnecessarily complex.	2.2 (1.64)
3. I thought the system was easy to use.	2.6 (1.14)
4. I think that I would need the support of a technical person to be able to use this system.	3.4 (0.55)
5. I found the various functions in this system were well integrated.	2.6 (0.55)
6. I thought there was too much inconsistency in this system.	3.0 (1.22)
7. I would imagine that most people would learn to use this system very quickly.	3.2 (0.84)
8. I found the system very cumbersome to use.	2.6 (1.52)
9. I felt very confident using the system.	3.0 (1.0)
10. I needed to learn a lot of things before I could get going with this system.	2.6 (1.67)
^a 1 = strongly disagree, 5=strongly agree	

Table 7: Iteration 1- Usability Sub-Scale Scores		
Sub-scale	SUS Item	Mean (SD)
Usability	2,3,5,6,7,8,9,10	2.73 (9.58)
Learnability	1,4	3.2 (1.25)

4.1.5 Age

The age of the participants in this sample varied. **Table 8** shows the different individual SUS scores in respect to the participants' ages.

Table 8: Iteration 1- Individual SUS Scores based on Age		
Subjects	Age	SUS Score
Subject1	42	65
Subject2	25	70
Subject3	43	82.5
Subject4	31	92.5
Subject5	60	42.5

The individual SUS scores did not show a relationship with age. It is evident that age was not a major factor affecting the satisfaction of the participants.

4.1.6 Session Elapsed Times

In each usability testing session, the participant was timed in completing all three tasks as a whole rather than each task separately. This method was taken because the three tasks were related to each other and simulate an activity that would be implemented in the workflow of a hemophilia-treating clinic. The time elapsed to complete the tasks was from the time the participant logged into the WAPPS-Hemo service until the end of the final task. For the first iteration, the five participants' time taken averaged 38.2

minutes. As shown in **Table 9**, it is worthy to note that Subject #5, who took the longest time to complete all tasks, had internet connection issues. The participant spent time checking and reconnecting to the internet, which resulted in increased time to complete the tasks.

Table 9: Iteration 1- Usability Session Times (minutes)	
Subject1	36
Subject2	31
Subject3	34
Subject4	42.5
Subject5	48

4.2 User Interface Modifications

Table 10 shows the modifications that have been implemented to the system by the programmers according to the usability testing results of the first iteration. While suggestions from usability testing were given the highest consideration, the final decision about implementation of the change were taken by the principal investigator, the author, the research coordinator and the programmers.

An important criterion considered in the decision was the frequency of any specific suggestions (i.e. suggestion with highest frequency were usually considered as

necessary); similarly, discrepant suggestions (i.e. modifying a function in opposite directions) were those more thoroughly considered as pointing out a possible problem, but not necessarily the proper solution to it. The check mark (✓) indicates changes that have been made, however, not necessarily as exactly as listed. The (✗) indicates the changes that were not implemented; explanations will be mentioned below.

Table 10: Iteration 1- Suggestions for Improvement and Implementations		
ID	Item	Modifications Implemented
1	<ul style="list-style-type: none"> • Remove the Add Measurement button • Add a Save button to the right of an active data point field (should disappear after the subject saves the respective entry) • Move the Finalize Measurements button to the centre, above the Request PK Calculation button • A pop-up message when a subject finalizes a measurement, including an OK and Cancel buttons • Request PK Calculation button could be greyed out 	<ul style="list-style-type: none"> ✓ ✓ ✓ ✓ ✗
2	<ul style="list-style-type: none"> • Replace the Add Infusion with a Save button • A pop-up message when subjects save their entries, including an OK and Cancel buttons • Add a title “Infusion Information” or “List of Available Infusions” 	<ul style="list-style-type: none"> ✓ ✓ ✓
3	<ul style="list-style-type: none"> • When subjects save their entries, the data point fields appear expanded and directly visible • Include a Collapse and Expand button 	<ul style="list-style-type: none"> ✓ ✓
4	<ul style="list-style-type: none"> • Add a button indicating “Select” or “Choose” button with a link 	<ul style="list-style-type: none"> ✗

	to the centre settings. (with small guide) <ul style="list-style-type: none"> • If not an admin, pop-up contacting Admin (name, email) 	✓
5	<ul style="list-style-type: none"> • Comment Rh not necessary 	✗
6	<ul style="list-style-type: none"> • Provide the units in parentheses (e.g. Infusion Duration (min)) 	✓
7	<ul style="list-style-type: none"> • A pop-up appears indicating that the data have been saved • A message under the Save button that the data have been saved 	✓ ✓
8	<ul style="list-style-type: none"> • Make warning signs bigger and/or add red warning text 	✓

There were several adjustments in item #1 that have been made that resulted with the highest number of occurrences among users. As seen in **Figure 8**, the Add Measurement button has been removed to lessen confusion users previously experienced. We added a *Plus* button instead of a Save button as suggested. However, the *Plus* button did the same intended function, which was to save respective measurements user entered. The *Plus* button has been placed to the right of the text fields to make it aligned and natural for users and thus provides easier usability. The Finalize Measurements button has been moved to the centre to make it more logical in sequence within the frame of the blood sample data entry fields. The Add an Infusion and Request PK Calculation buttons have been placed further at the bottom to appear separate from the blood sample measurement data fields. The Request PK Calculation was not greyed out as suggested because it would require a lot of programming and underlying modifications. It was thought that after the aforementioned modifications were implemented, it will not remain

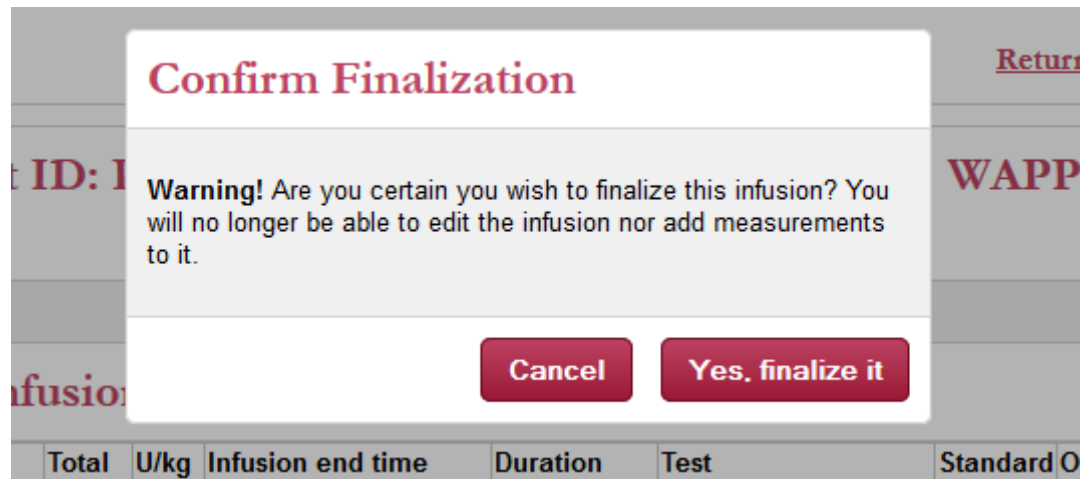
as a user interface obstacle. When users finalized their measurements, there was a pop-up message confirming their data entry will be finalized and cannot be edited. In addition, a Cancel and Yes buttons have been added to give users reassurance and to re-think their decisions if needed, **Figure 9**.

Figure 8 Modification of Data Entry Buttons

Collected Infusion Data

+/-	ID	Drug	BW (kg)	Total U	U/kg	Infusion end time	Duration (mins)	Test	Standard	Opt.	Notes	Tools
●	267	Advate	80	4000	50.0	2012-06-01 08:00 AM	0	One Stage Coag. (PTT Based)	Generic	☰		
Measurement Date/Time			Time Elapsed (h:m)			Special Measure		Concentration Notes				
<input type="text" value="2012-06-01"/> <input type="text" value="08:00 AM"/> ⬇️ ⬆️ ⬇️ ⬆️			TBD			<input type="checkbox"/> Pre-dose <input type="checkbox"/> Below Threshold		<input type="text"/> <div style="border: 1px solid #ccc; height: 40px; width: 100%;"></div> +				
<div style="background-color: #800000; color: white; padding: 5px 20px; display: inline-block; border-radius: 5px;">Finalize Measurements</div>												
<div style="display: flex; justify-content: space-around; margin-top: 10px;"> <div style="background-color: #800000; color: white; padding: 5px 15px; display: inline-block; border-radius: 5px;">Add an Infusion</div> <div style="background-color: #800000; color: white; padding: 5px 15px; display: inline-block; border-radius: 5px;">Request PK Calculation</div> </div>												

Figure 9 Confirmation Message



In item #2, the Add Infusion button under sub-title Recommended Fields confused users. Alternatively, the Add Infusion button was replaced with Save Infusion button and sub-title Recommended Fields was replaced with Optional Fields, because the semantics was an issue for several subjects. **Figure 10** shows that when users enter new infusion data, a window with text fields pop-up, called a splash screen. The purpose of the splash screen was to help users focus on one infusion data entry at a time. Included are Cancel and Save Infusion buttons, this is to make it easier for users to know and understand that their data have been saved or not. A confirmation or warning message was needed at that stage because users cannot edit that information at a later time. This enhancement has also remedied interface problem #7. A minor change has been made to the sub-title to “Collected Infusion Data” to help users understand which section they are working on.

Figure 10 Infusion Data Entry

Infusion Data

Required Fields

Drug: Advate

Lab Test: One Stage Coag. (PTT Based)

Standard: Generic

Body Weight: (kg) **Required**

Total Units Administered: (IU) **Required**

Units per kg: 50.00 (IU/kg)

End of Infusion Date: **Required**

End of Infusion Time: **Required**

Infusion Duration: (min) **Required**

Optional Fields

Height: (cm)

Hematocrit: (%)

Hemoglobin:

Serum Creatinine: (µmol/L)

vWF:RiCof: (µmol/L)

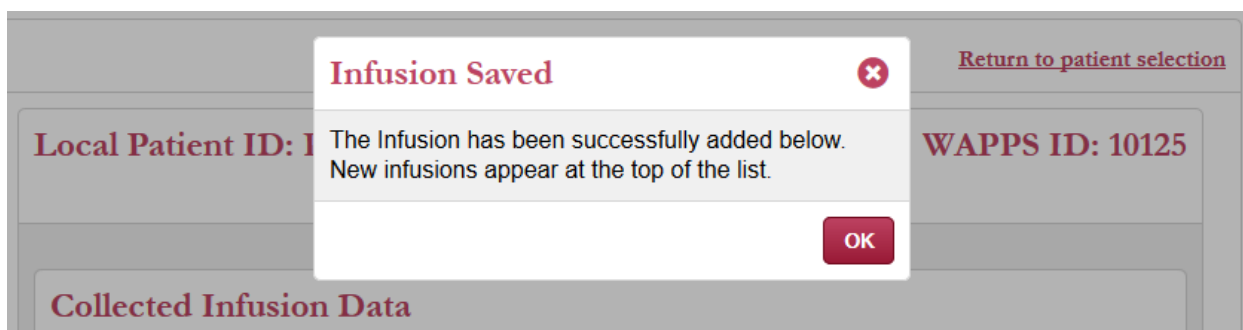
vWF:Ag: (µmol/L)

Notes:

Cancel Save Infusion

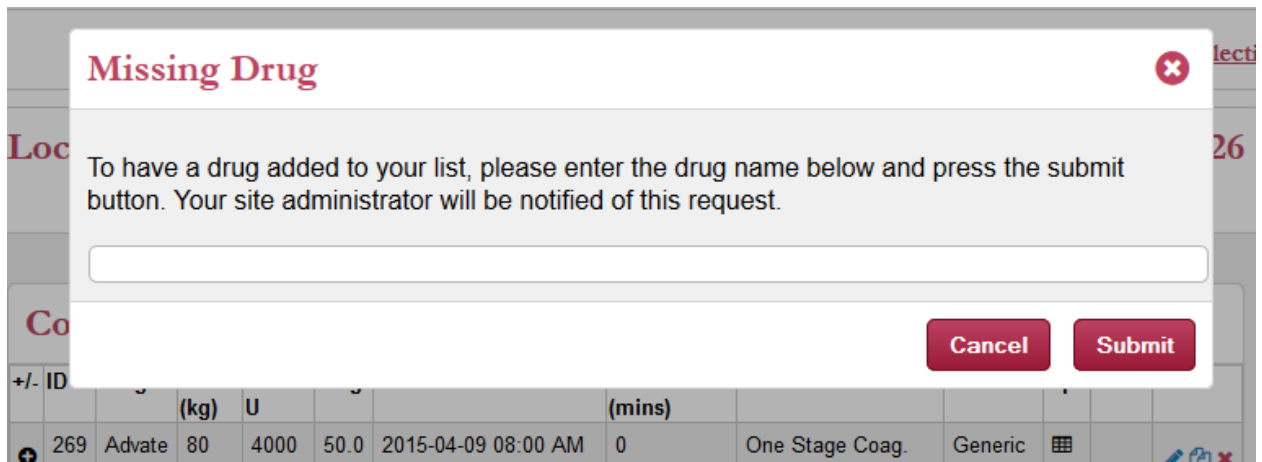
Item #3 was fixed by making the entire section appear when users added a first infusion data of a patient. Additionally, a pop-up message was added to help users confirm that the infusion data was successfully stored, as seen in **Figure 11**.

Figure 11 Infusion Confirmation



For item #4, the programmers remedied the interface problem slightly differently from the suggestion stated, however, this change should suffice to achieve the same goal of that particular function. If users wanted to select a drug that was not on the list they had the option to click the Missing Drug link next to the Drug field, as shown in **Figure 10**. The user will be presented with dialog box asking for the required drug's name and will be added by the appropriate personnel, as seen in **Figure 12**. Item #5 was modified by adding an “info” icon notifying users that it is not necessary to indicate whether the blood type is negative or positive. **Figure 10** shows that interface problem # 6 was easily remedied by simply adding the text (*min*) after the text field. Interface problem #8 was resolved by adding “Required” text in red next to the warning symbol as seen in **Figure 10**, as well.

Figure 12 Missing Drug Addition



Technical Problems		
1	After adding an infusion, subjects are frustrated they cannot go back to edit a mistake in patient profile	✓
2	Subject was frustrated that Infusion Duration would not accept numbers with decimals (e.g. 5.5)	✗

Technical Problems

As mentioned previously, technical problems are usually not related to user interface obstacles and experiences. Item #1 has been partially improved as seen in **Figure 13**. The programmers made patients’ profiles editable, however, an attempt was made to identify that certain items cannot be edited, noted by “Static Fields” and other items that are editable, noted by “Editable Fields”. For item #2, in the Infusion Duration field, when a user entered a number with a decimal point, the system will consider the

entry invalid. Consequently, users were required to enter 0 minutes if the duration was less than 10 minutes, where, practically, infusion durations were less than 10 minutes.

Figure 13 Edit Patient Panel

Edit Patient

Static Fields

Gender: Male
Date of Birth: 1997-04-10
Local Patient ID: Px3

Editable Fields

Blood Group: B
Baseline Factor Level: [input field] (U/ml)
Positive History of Inhibitors:

Cancel Save

Px3 18 M B [edit icon]

4.3 Second Iteration

In the second iteration, eight participants were tested over a period of three weeks.

4.3.1 Pre-test Background and Demographics: Second Iteration

The background questionnaire that was done prior the usability testing showed that one participant was a physician and/or a hematologist (12.5%), two were registered nurses or nurse practitioners (25%), two administrators (25%), and three indicated as “other” (37.5%), more specifically a clinical research assistant, a research coordinator and

a data manager. Additionally, two participants (25%) were using a PK calculator at the time of the testing and three participants had prior experience with the WAPPS-Hemo system. Experience ranged from two weeks to six months. This is the same questionnaire that was used in the first iteration. **(Table 11)**

Table 11: Iteration 2- Demographic Descriptors of Participants (n=8)			
		n	%
Gender	Male	2	25
	Female	6	75
Age (years)	Mean	44	
Role in Treating Centre	Physician/Hematologist	1	12.5
	RN/Nurse Practitioner	2	25
	Administrator	2	25
	Other (clinical research assistant, research coordinator, data manager)	3	37.5
Experience	Use of any PK calculator	2	25
	WAPPS-Hemo	3	37.5

4.3.2 Usability

A list of user interface problems identified by the eight participants is presented in **Table 12**. This list identifies the interface design items that subjects had trouble with pertaining to the number of occurrences after modifying and enhancing the system according to the findings of the first iteration. The higher the number of occurrences were, the higher the priority given to attempt changes by the WAPPS-Hemo programmers. Again, interface problems were defined as user interface and design issues and misconceptions. Technical problems were defined as issues that appear to users due to back-end (server side) restrictions. These issues may be considered long-term problems and their solutions could come at a later stage in the development of the system.

Table 12: Iteration 2- Usability Issues		
Interface Problems		
ID	Item	Number of Occurrences
1	The Special Measure section is not understood, especially the definitions of “Pre-dose” and “Below Threshold”.	3
2	The “plus” button does not appear sufficiently explain saving the blood sample measurement.	3
3	The Measurement Date/Time fields does not intuitively indicate that “lab work” must be entered.	2
4	The coloring scheme is not consistent on different browsers, such as IE, Safari and Chrome	1
5	User did not understand why the Copy icon/function existed, even after reading the pop-up message after finalizing measurement.	1
6	The Save Infusion and Cancel buttons are visible in small resolution displays	1
7	The header row in the Patient List table has small fonts and small arrow icons	1
8	Difficult to understand the steps of Adding and Activating Drugs	1
9	Results page was confusing and unnecessarily cluttered	1
Technical Problems		
1	After adding an infusion, subjects are frustrated they cannot go back to edit a mistake in the infusion profile	3

As mentioned before, suggestions for improvements were coded (**Appendix H**). During the process of thinking out loud, subjects gave their opinions and visions on how certain aspects or functionality of the WAPPS-Hemo could have been enhanced. These comments were identified under the suggestions for improvements category. The author applied the concurrent and retrospective probing techniques where subjects had the opportunity to express any other ideas to improve the development of the system during the testing after they had finished the required tasks.

4.3.3 Interface Problems

Three out of eight participants had difficulty understanding the Special Measure section of the service when they entered the data for the blood sample draws, see **Figure 14**. This was one of the two highest occurring problems with the new website interface design. When participants reached that section, they paused and tried to understand what Pre-dose and Below Threshold meant. One participant commented, “I have no idea what to do with this special measure part”. Despite hovering over the “info” icons (the blue circle with the letter “i”), participants were not satisfied with the information provided. This may have been a problem for non-clinicians. A nurse also had trouble as he or she commented, “[This] is not intuitive to me as a nurse, it's not clear”. The “info” icons provided a non-detailed explanation of what Pre-dose and Below Threshold meant and the participants still felt overwhelmed. Therefore, participants skipped this section altogether.

Figure 14 Sample Data Entry New Design

+/-	ID	Drug	BW (kg)	Total U	U/kg	Infusion end time	Duration (mins)	Test	Standard	Opt.	Notes	Tools
●	293	Advate	68	3000	44.1	2015-05-04 12:00 AM	5	One Stage Coag. (PTT Based)	Generic	☰		
Measurement Date/Time			Time Elapsed (h:m)		Special Measure		Concentration	Notes				
<input type="text" value="2015-05-04"/> <input type="text" value="12:00 AM"/>			TBD		<input type="checkbox"/> Pre-dose <input type="checkbox"/> Below Threshold		<input type="text"/>	<input type="text"/> <input type="text"/> <input type="text"/>				

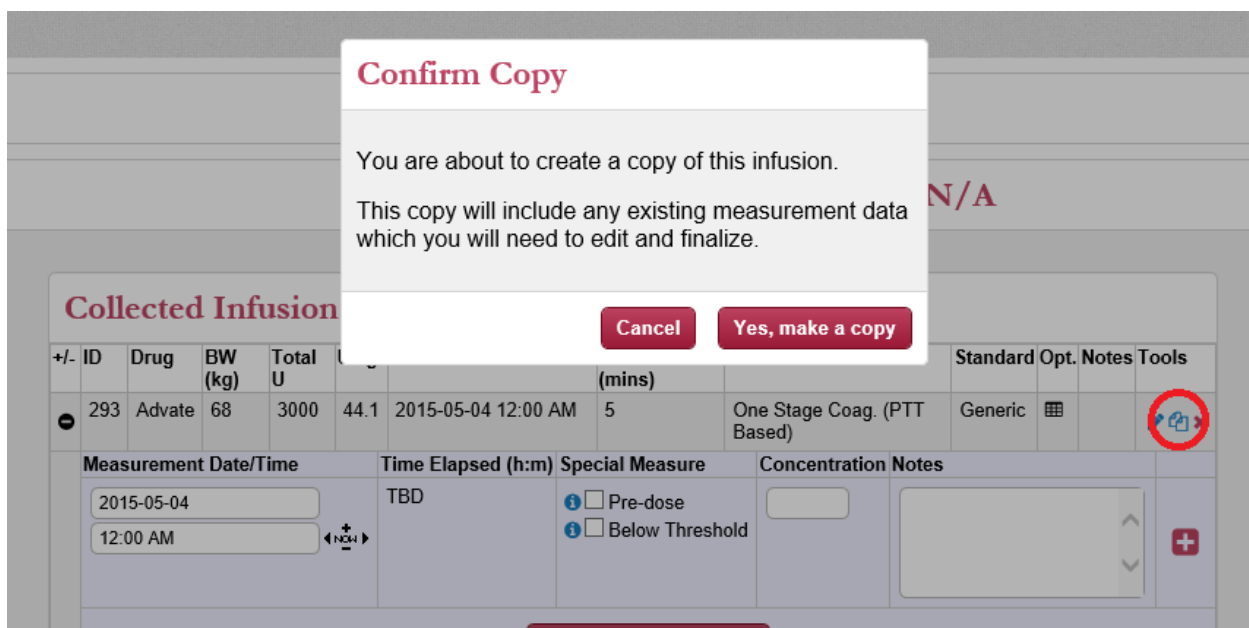
Another interface design hindrance that arose with the same number of occurrences was the *Plus* button that was intended to be used to add different blood sample draw entries, as seen in **Figure 14**. Participants thought that the *Plus* button was meant to add more additional infusions and not the blood sample draw data. Generally, users wanted to “save” any data entry they did on the system, but this step was not performed adequately during the testing. One participant commented, “I would naturally want to hit something else to ensure that it's saved...”. That participant was looking for a *save* button or a clue to what to do after entering the blood sample draw information. Another participant commented after several attempts, “I'd like to see this button more intuitive because I overlooked it”. Clicking on a “plus” button might have been a spontaneous action, nevertheless, these participants' experience proved the opposite response.

A third major interface issue two participants found was the “Measurement Date/Time” subtitle. When users were ready to enter blood sample draw data, they initially did not know where to put the data. Semantics was the problem in this situation.

One participant was confused and commented, “not the screen I wanted, I want a screen that says lab results”. This showed that participants might have experience with other systems that they immediately looked for wording or language that they were familiar with.

The following are six different interface design issues occurred to six different individuals (one occurrence each). First, a participant commented that the color schemes were different on different web browsers and operating systems. For example, color schemes seemed altered in Google Chrome compared to Mozilla Firefox or Internet Explorer, in addition to a difference in Safari under Mac OSX. Another participant pointed out the Copy icon as shown in **Figure 15**. The participant did not know why that function existed, even after they have clicked on the icon and read the pop-up message. The user noted in frustration, “I’m not sure what does the copy button does exactly”. The participant did not fully understand the purpose of the message.

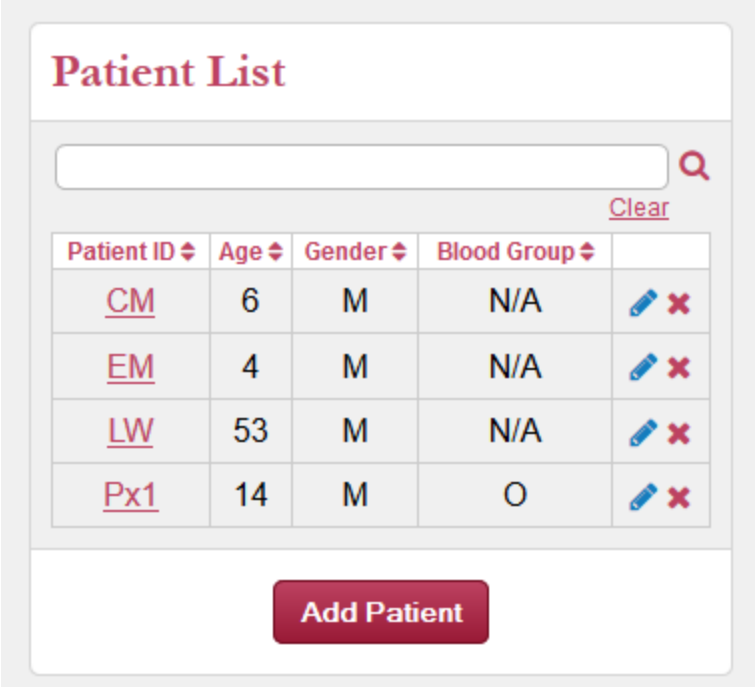
Figure 15 Copy Confirmation



Users with small display or poor screen resolutions had difficulty seeing the “Save Infusion” and “Cancel” button when they input infusion data. A participant was unhappy

because the splash screen could not be dragged or moved to make these vital buttons visible. Hence, the participant clicked the “X” button to close the splash screen, which lost the information entered, **Figure 10**. The header row in the Patient List table, **Figure 16**, was small, which made it difficult for the participant to notice the arrow icons used for sorting the order of the list. The user commented on the small size of the font and scroll icons.

Figure 16 Patient List



The screenshot shows a web interface titled "Patient List". At the top, there is a search bar with a magnifying glass icon and a "Clear" link. Below the search bar is a table with four columns: "Patient ID", "Age", "Gender", and "Blood Group". Each column header has a small downward arrow icon. The table contains four rows of patient data. Each row has a blue pencil icon and a red 'X' icon in the fifth column. Below the table is a red button labeled "Add Patient".

Patient ID	Age	Gender	Blood Group	
CM	6	M	N/A	
EM	4	M	N/A	
LW	53	M	N/A	
Px1	14	M	O	

One participant’s hemophilia centre did not have the appropriate drug preloaded, Therefore, they had to add it manually. Even though the user clicked the hyperlink to help add the drug, “Missing Drug?” as seen in **Figure 12**, the user was confused because there was no indication to what the next steps were. At that point, the participant needed

guidance from the observer. Although the user obtained some help with the navigation, the sequence and functionality of adding a drug was not coherent or clear to the user. He or she concluded with a “that's not very clear, you need to fix that” comment. Finally, the same participant had trouble understanding the layout of the Results page. He or she thought that it was cluttered and too busy and presented no useful information **Figure 17**.

Figure 17 Patient Results Page

Patient Results										
Return to patient list										
Approved										
Submitted	Patient Id	Age	Blood Group	Inhibitor	Compartment Data	Half-life (hr)	Time to 0.05 IU/ml (hr)	Time to 0.02 IU/ml (hr)	Time to 0.01 IU/ml (hr)	Notes
Rejected										
Submitted	Patient Id	Age	Blood Group	Inhibitor	Compartment Data	Half-life (hr)	Time to 0.05 IU/ml (hr)	Time to 0.02 IU/ml (hr)	Time to 0.01 IU/ml (hr)	Notes
Awaiting Approval										
Submitted	Patient Id	Age	Blood Group	Inhibitor	Compartment Data	Half-life (hr)	Time to 0.05 IU/ml (hr)	Time to 0.02 IU/ml (hr)	Time to 0.01 IU/ml (hr)	Notes

In regards to technical problems, only one issue was encountered by three people. The problem was that after adding an infusion and subsequent blood sample draw data, the participant was not able to edit that information after finalizing the measurements. This was a recurrent problem because it occurred in the first iteration. It is yet unknown whether this feature will be implemented into the current development stage of the WAPPS-Hemo system.

Table 13 shows a list of the improvements suggested by the participants after the completion of the second iteration. Similar to the first iteration, the suggestions came mainly from the usability interviews. However, they might need to be slightly modified to be improved by the programmers according to the feasibility and appropriateness of the suggestion.

Note: ID corresponds to the issue item in **Table 12**.

Table 13: Iteration 2- Suggestions for Improvements of the User Interface	
ID	Item
1	Provide more detailed explanations and definitions to Pre-dose and Below Threshold
2	Replace the “plus” sign with the word “Save” or “Save and Add”
3	Replace Measurement Date/Time with “Lab Work” and Concentration could be replaced with “Factor Concentration”
4	Provide more information on the purpose of the Copy feature. A reference to it might be added in the warning of Finalizing Measurement.
5	Provide more flexible date and time formatting (Canadian vs US)
6	Add an “info” icon next to Finalize Measurement instead of hover pop-up
7	Replace +/- icon with a title, tip or a more resemblance icon
8	Add a hyperlink to a guide for Add Drug feature
9	De-clutter, only show what’s actually available (approved, rejected or pending)
10	“Return to patient selection” could be bigger with a “back” arrow
11	System could accept % for concentration as some concentrations are in percent (US)
12	Add a pre-populated Drug List
13	“Info” icons can be added to Standard: Generic or Specific field with appropriate definitions and explanations
14	“Info” icons can be added to Concentrations to explain exactly what to be entered and

	what units (some centres have labs results come in percentages)
15	Auto-populate Concentrations field with “0” (i.e. “0.xxx”) or automatically generate decimal points upon entering whole number
16	Be able to hit the Enter key on keyboard to complete the Concentration entry
17	Add a message to inform user about the 24 hr time frame to receive a PK result upon requesting one
18	Combined Data Entry and Results tab in to 1 called Patients, then add a 2 columns to the Patient List, one for Patient Entry link and the for Results status: available, pending, rejected

4.3.4 SUS

As aforementioned, five non-experienced participants were included to calculate the SUS score. Participants were given the SUS survey immediately after completion of the usability testing. The score achieved for this iteration was 73 out of 100. The overall SUS score increased from 70.5 to 73, which is a 2.5% increase. The average scores and standard deviations (SD) were calculated for each SUS item shown in **Table 14**.

Table 14: Iteration 2- System Usability Scale Mean Scores	
	Subjects (n=5) Mean (SD)^a
1. I think that I would like to use this system frequently.	3.0 (0.71)
2. I found the system unnecessarily complex.	3.2 (1.30)
3. I thought the system was easy to use.	3.0 (1.22)
4. I think that I would need the support of a technical person to be able to use this system.	3.0 (1.58)
5. I found the various functions in this system were well integrated.	2.8 (1.30)
6. I thought there was too much inconsistency in this system.	3.2 (1.30)
7. I would imagine that most people would learn to use this system very quickly.	2.8 (1.30)
8. I found the system very cumbersome to use.	3.2 (1.10)
9. I felt very confident using the system.	2.6 (1.14)
10. I needed to learn a lot of things before I could get going with this system.	2.4 (1.52)
^a 1 = strongly disagree, 5=strongly agree	

Table 15 shows the usability and learnability sub-scale scores.

Table 15: Iteration 2- SUS Sub-scale Scores		
Sub-scale	SUS Item	Mean (SD)
Usability	2,3,5,6,7,8,9,10	2.92 (9.78)
Learnability	1,4	2.70 (1.98)

4.3.5 Age

The age of the participants in this sample varied. **Table 16** shows the different individual SUS scores respective to the participants' ages.

Table 16: Iteration 2- Individual SUS Scores based on Age		
Subjects	Age	SUS Score
Subject1	35	100
Subject2	57	77.5
Subject3	47	92.5
Subject4	58	50
Subject5	46	40

The individual SUS scores did not show a relationship with age. It is evident that age was not a major factor affecting the satisfaction of the participants

4.3.6 Session Elapsed Times

As it was done in the first iteration, participants were timed in completing all three tasks as a whole rather than each task separately. The time elapsed to complete the tasks was from the time the participant logged in to the WAPPS-Hemo website until the end of the final task, shown in **Table 17**. For this iteration, the five participants with no previous exposure to the system averaged at 41.2 minutes.

Table 17: Iteration 2- Usability Session Times (in minutes)	
Subject1	57
Subject2	51
Subject3	32
Subject4	29
Subject5	37

4.3.7 Experienced Users

The following shows the results of the three participants that were excluded from the analysis of the second iteration, due to their previous experience with the WAPPS-Hemo service. Their SUS score was 74.2 out of 100.

Table 18: Iteration 2- SUS Scores of Experienced Users	
	Subjects (n=3) Mean (SD)^a
1. I think that I would like to use this system frequently.	3.7 (0.58)
2. I found the system unnecessarily complex.	2.7 (0.58)
3. I thought the system was easy to use.	2.7 (1.15)
4. I think that I would need the support of a technical person to be able to use this system.	2.7 (1.53)
5. I found the various functions in this system were well integrated.	3.0 (1.00)
6. I thought there was too much inconsistency in this system.	2.7 (1.53)
7. I would imagine that most people would learn to use this system very quickly.	2.3 (1.53)
8. I found the system very cumbersome to use.	3.0 (1.00)
9. I felt very confident using the system.	3.3 (0.58)
10. I needed to learn a lot of things before I could get going with this system.	3.7 (0.58)
^a 1 = strongly disagree, 5=strongly agree	

Table 19 shows the usability and learnability sub-scale scores.

Table 19: Iteration 2- SUS Sub-scale Scores Experienced Users		
Sub-scale	SUS Item	Mean (SD)
Usability	2,3,5,6,7,8,9,10	2.92 (7.94)
Learnability	1,4	3.20 (2.10)

Table 20: Iteration 2- Individual SUS Scores based on Age (Experienced Users)		
Subjects	Age	SUS Score
Subject1	26	72.5
Subject2	58	55
Subject3	--	97.5

Table 21: Iteration 2- Usability Session Times (in minutes) of Experienced Users		
		Average Time
Subject1	37	40.7
Subject2	70	
Subject3	15	

4.4 Summary of Results

The following table shows the SUS scores achieved from both iterations. The scores remained above average and in the category “good”.

Table 22: Summary of SUS Scores

	Iteration 1	Iteration 2
SUS Score	70.5	73

CHAPTER 5: DISCUSSION AND FUTURE WORK

5.1 Discussion

The study presented entailed the usability testing and evaluation of the WAPPS-Hemo web application created in the HIRU at McMaster University. The web application is intended to help calculate PK assessments of factors VIII and IX for the treatment of patients with hemophilia. We wanted to know what health care professionals encounter when they used the system for the first time. The objective was to collect participants' feedback, insights and ideas on how we could improve the system's usability and user interface through its development stages.

The methods used to conduct the study were the think aloud technique (Monique W M Jaspers et al., 2004; Rubin & Chisnell, 2008) and the SUS (Brooke, 1996; "System Usability Scale (SUS)," 2013) through two iterative usability testing sessions. The think aloud technique has been chosen for this type of testing because of its effectiveness and usefulness shown in other usability testing of software applications in their development cycles (Kobak et al., 2011; Rubin & Chisnell, 2008; J. Stinson et al., 2014; Wakefield et al., 2015). These methods provided valuable and useful input in identifying major user interface design problems encountered while using the WAPPS-Hemo service.

During the usability testing sessions, the think aloud technique helped us identify different problems a user would potentially encounter during their interaction with the system via comments and feedback given by the participants. The majority of the issues identified were heavily related user interface design problems and, to a lesser extent,

technical problems. Most of the issues that were identified during the first iteration were remedied for second iteration of testing the WAPPS-Hemo. Even though most participants expressed frustrations when they interacted with the system, they shared positive comments about the system. Quantitative data were also collected and analyzed, where SUS scores and session durations were calculated.

After consolidating all the qualitative data and field notes, we found 88 instances of user design interface problems classified in the first iteration of the usability testing. In the second iteration, however, there were only 14, which equates to an 84.1% reduction in the instances of interface problems. This reduction is attributed to the changes and improvements implemented after the first iteration. We also saw better overall user satisfaction. In regards to user satisfaction, the scores were assessed in both iterations where any SUS score above 68 is considered above average (Brooke, 1996).

The SUS score for the first iteration was 70.5, while the second iteration of the usability testing achieved 73. This is a 2.5% increase of the user satisfaction scale after testing the enhanced version of the WAPPS-Hemo service, especially being in its early development phase. It was expected to gain better general satisfaction from participants because we have repaired the problems that arose and improved the development of the service. We did not have major discrepancies between the inclusion criteria of participants in both iterations. The relevant roles of participants were almost identical except that we had two physicians and one nurse in the first iteration while we had two nurses and one physician in the second. Selection of participants with different roles was intentional to simulate a real-world scenario by whom the WAPPS-Hemo can be used.

Additionally, the SUS scores of the individual participants did not show a clear relationship between the scores and their ages. Possibly a larger number of participants in each sample could provide more data points that could help us understand an association between usability and age. The SUS sub-scale refers to the ease of usability and learnability with scores ranging from 0 to 4, where 0 being the lowest and 4 the highest score (Bangor et al., 2009). The average usability score in the SUS sub-scale in the first iteration was 2.73 and 2.92 in the second iteration. The average learnability was 3.2 in the first iteration and 2.7 in the second iteration. The SUS sub-scale *usability* refers to how easy and pleasant users found the system interface design, and *learnability* is how long it takes users to complete specific tasks in the system (Nielsen, 2012). The *usability* and *learnability* scores could indicate that the WAPPS-Hemo was easier to use and quicker to learn in the second iteration. However, statistical significance between both iterations was not computed due to the small sample size. In addition, we received positive comments on the cleanliness and simplicity of the web application. Other participants' comments revolved around how easy was the system to learn after inputting the first infusion data of a patient.

The number of participants in the sample is relatively small for providing a more significant representation and interpretation of the results. Another reason might be that new interface problems have occurred in the second iteration, therefore, the subjective perceptions by the participants played a main role in the new interface design's learnability.

The SUS score of the three experienced users with the WAPPS-Hemo was 74.2. However, since no statistical analysis was conducted to see if there is a statistical significant difference between this score and the score of the previous iteration, no definite conclusion can be made on the improvement of the usability of the system at different iterations. During the usability testing sessions, the observer noticed that these participants confidently navigated through the web application. This could indicate that when users have previous exposure to the WAPPS-Hemo service, the level of frustration can be less.

The average time durations the participants completed all three tasks for the first and second iterations were 38.2 minutes and 41.2 minutes, respectively. In general, the total time duration is expected to remain the same or decrease. One reason that could explain the potential time increase is that Concurrent Probing was used during the sessions that caused the increased time spent on the tasks during the second iteration. Additionally, intermittent network problems and uncontrollable server errors occurred during the usability testing.

Based on feedback from the current usability testing, the WAPPS-Hemo program has been launched in June 2015 and is currently employed by over 20 centres and 100 users.

5.2 Limitations

There were challenges and limitations to this study. One of the limitations was we did not find in the literature an SUS baseline score specific to applications in the PK field to compare our initial score with. A meta-analysis of more than 3,500 SUS results across multiple fields of applications of tested software showed that the SUS can provide more information than just a scale (Brooke, 2013). As aforementioned, an average score is 68, but there are other attributes to the scale, such as “acceptability ranges” and “adjective ratings”. According to the meta-analysis, a score range from 68-84 is considered “good” and “acceptable” (Bangor et al., 2009; Brooke, 2013).

Another limitation was the number of iterative cycles of the usability testing, which was two only in this case. In general, there is no limit to the number of iterative cycles a software in development can go through. Typically, the development involves an initial cycle followed by evaluation and feedback on the system design, and finally leading further cycles until an acceptable system evolves (A. Kushniruk, 2002). Many of the works we looked at involved 1-4 cycles in system design, we chose two iterative cycles due to time constraints. Moreover, since we did not observe a dramatic change in the SUS score between before and after the process, and since we started with a high baseline SUS, we are confident in relying on constructive feedback from experienced real-world users of the system for future improvements rather than further cycles of usability testing at this stage.

Second, a small number of participants were in each session. Albeit this small number could produce adequate qualitative data to conduct the think aloud technique, a larger sample is more desirable in conducting quantitative approaches such as the SUS.

We recruited eight participants for the second iteration, three had been excluded because of their prior experience with the WAPPS-Hemo service, which could have created biased results regarding the SUS score. The background questionnaires and the SUS surveys were sent to the participants beforehand via email, but they were only asked to complete the background questionnaire before the usability test. However, it is possible that some participants have looked at the SUS survey questions before the testing. This, as a result, could have steered their way of thinking and their expectations.

The conventional method to conduct usability testing would be to have participants physically present with the researcher. One of the advantages of that approach over remote testing is that the researcher would be able to observe the participants while using the WAPPS-Hemo and be able to record their facial expressions and body gestures. These indicators would help further analyze different aspects of obstacles that participants encounter. However, with remote testing, the researcher had to rely entirely on transcriptions and vocal tones from the recording.

Participants had the opportunity to use the WAPPS-Hemo service only once during the usability testing. They have not seen or interacted with the system after the suggested improvements were implemented. This approach was suitable to compare the SUS scores of the two iterations. However, it might have also been useful to collect and

analyze experienced users' comments and feedback after changes to the system have been implemented. In other words, it is useful to have the same users of the first iteration participate again in the second iteration to compare their feedback and comments and see if they think the problems they faced in the system the first time have been addressed. Some subjects mentioned that if they have interacted with the system more than once, they would have been very familiar with it, its features, and their way around the system.

The time spent on individual tasks was not measured. This could have given us an indication on how much time users spent on particular areas on the website or specific tasks as proposed in the Wakefield study (Wakefield et al., 2015). Rather we have measured the timing of the session as a whole, which did not provide us with representative results of each individual task. Additionally, observer involvement was also a factor in the total time spent, since subjects would stop and ask questions at certain points while using the system.

5.3 Implications and Future Work

There are several ways that we can expand this study. One way would be to increase the number of usability testing iterations. Having more iterations and possibly increased number of participants in each round could provide us with more understanding of how people perceive the WAPPS-Hemo service. These iterations can incorporate the think aloud technique to gain qualitative feedback from users with experience interacting

with the WAPPS-Hemo. In addition, to gain a significant quantitative outcome, we can send the SUS survey to a larger sample of people.

A larger sample would also facilitate the use of statistical analysis, which we plan to conduct at a later stage with users trained in PK, to try to selectively evaluate the impact of the software and not of the required basic knowledge. Alternatively, we will consider a dedicated training of tester on population PK before the testing sessions.

In addition, since we only measured the time spent during an entire session, it would be useful to also measure the time elapsed with each task during the think aloud technique. This method could help us stratify the difficulties potential end users might encounter while performing specific tasks on the system. We could also calculate the number of times participants asked for assistance from the tester. Additionally, the number of material errors occurred during performing the task of the testing exercise could be calculated to help us indicate areas in need for more clarifications or changes.

Many participants had difficulty understanding the Pre-dose and Below Threshold options upon inputting infusion blood sample draws. Nevertheless, we could modify the fabricated patient data to allow participants experiment to allow us obtain further explanation on how they perceived these specific options. Henceforth, we could implement extra features to facilitate these functions.

It is worthy to mention that the transcriptions and qualitative analyses were performed by the author only. This could have biased determining the different themes

and categories of interface problems and usability issues. At least two testers could have contributed in the decision of these themes and analysis to avoid such bias.

CHAPTER 6: CONCLUSION

We conducted usability testing and analyses to determine the usability of the WAPPS-Hemo application among various potential users. Iterative usability testing sessions showed improvements in user overall satisfaction and SUS score, comparing the modified interface design to the original design of the system. This application is ready for a clinical study assessing utility in practice, but should continue to incorporate end-user feedback on its design.

REFERENCES

- Aledort, L. M., Haschmeyer, R. H., & Pettersson, H. (1994). A longitudinal study of orthopaedic outcomes for severe factor-VIII-deficient haemophiliacs. *Journal of Internal Medicine*, *236*, 391–399. <http://doi.org/doi:10.1111/j.1365-2796.1994.tb00815.x>
- Arsand, E., & Demiris, G. (2008). User-centered methods for designing patient-centric self-help tools. *Informatics for Health & Social Care*, *33*, 158–69. <http://doi.org/http://doi.org/10.1080/17538150802457562>
- Bangor, A., Kortum, P., & Miller, J. (2009). Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of Usability Studies*, *4*, 114–123. <http://doi.org/66.39.39.113>
- Björkman, S. (2010). Limited blood sampling for pharmacokinetic dose tailoring of FVIII in the prophylactic treatment of haemophilia A. *Haemophilia*, *16*, 597–605. <http://doi.org/10.1111/j.1365-2516.2009.02191.x>
- Borsci, S., Federici, S., & Lauriola, M. (2009). On the dimensionality of the System Usability Scale: A test of alternative measurement models. *Cognitive Processing*, *10*, 193–197. <http://doi.org/10.1007/s10339-009-0268-9>
- Breakey, V. R., Warias, A. V, Ignas, D. M., White, M., Blanchette, V. S., & Stinson, J. N. (2013). The value of usability testing for Internet-based adolescent self-management interventions: “Managing Hemophilia Online”. *BMC Medical Informatics and Decision Making*, *13*, 1–14. <http://doi.org/10.1186/1472-6947-13-113>
- Britto, M. T., Jimison, H. B., Munafò, J. K., Wissman, J., Rogers, M. L., & Hersh, W. (2009). Usability Testing Finds Problems for Novice Users of Pediatric Portals. *Journal of the American Medical Informatics Association*, *16*, 660–669. <http://doi.org/10.1197/jamia.M3154>
- Brooke, J. (1996). *SUS: A “quick and dirty” usability scale. Usability Evaluation in Industry*. London, England: Taylor & Francis.
- Brooke, J. (2013). SUS : A Retrospective. *Journal of Usability Studies*, *8*, 29–40. Retrieved from http://www.usabilityprofessionals.org/upa_publications/jus/2013february/brooke1.html
<http://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html>

- Collins, P. W., Fischer, K., Morfini, M., Blanchette, V. S., & Björkman, S. (2011). Implications of coagulation factor VIII and IX pharmacokinetics in the prophylactic treatment of haemophilia. *Haemophilia*, *17*, 2–10. <http://doi.org/10.1111/j.1365-2516.2010.02370.x>
- Cristancho-Lacroix, V., Moulin, F., Wrobel, J., Batrancourt, B., Plichart, M., De Rotrou, J., ... Rigaud, A.-S. (2014). A web-based program for informal caregivers of persons with Alzheimer's disease: an iterative user-centered design. *JMIR Research Protocols*, *3*(3), e46. <http://doi.org/10.2196/resprot.3607>
- Daniels, J., Fels, S., Kushniruk, A., Lim, J., & Ansermino, J. M. (2007). A framework for evaluating usability of clinical monitoring technology. *Journal of Clinical Monitoring and Computing*, *21*, 323–330. <http://doi.org/10.1007/s10877-007-9091-y>
- Factor Replacement Therapy. (2015). Retrieved August 11, 2015, from <http://www.hemophilia.ca/en/bleeding-disorders/hemophilia-a-and-b/the-treatment-of-hemophilia/factor-replacement-therapy/>
- Graham, T. a D., Kushniruk, A. W., Bullard, M. J., Holroyd, B. R., Meurer, D. P., & Rowe, B. H. (2008). How usability of a web-based clinical decision support system has the potential to contribute to adverse medical events. *AMIA Annual Symposium Proceedings*, 257–261.
- Gringeri, A., Lundin, B., Von Mackensen, S., Mantovani, L., & Mannucci, P. M. (2011). A randomized clinical trial of prophylaxis in children with hemophilia A (the ESPRIT Study). *Journal of Thrombosis and Haemostasis*, *9*, 700–710. <http://doi.org/10.1111/j.1538-7836.2011.04214.x>
- Hsieh, H.-F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, *15*, 1277–1288. <http://doi.org/10.1177/1049732305276687>
- Iorio, A. (2014). WAPPS-Hemo (Web Accessible Population Pharmacokinetics Service - Hemophilia) Protocol – Phase 1: Development and Implementation. Health Information Research Unit (HIRU), McMaster University.
- Jaspers, M. W. M. (2009). A comparison of usability methods for testing interactive health technologies: Methodological aspects and empirical evidence. *International Journal of Medical Informatics*, *78*, 340–353. <http://doi.org/10.1016/j.ijmedinf.2008.10.002>
- Jaspers, M. W. M., Steen, T., van den Bos, C., & Geenen, M. (2004). The think aloud method: a guide to user interface design. *International Journal of Medical Informatics*, *73*, 781–95. <http://doi.org/10.1016/j.ijmedinf.2004.08.003>

- Johnson, C. M., Johnson, T. R., & Zhang, J. (2005). A user-centered framework for redesigning health care interfaces. *Journal of Biomedical Informatics*, 38(1), 75–87. <http://doi.org/10.1016/j.jbi.2004.11.005>
- Khajouei, R., & Jasper, M. W. M. (2010). The impact of CPOE medication systems' design aspects on usability, workflow and medication orders: a systematic review. *Methods of Information in Medicine*, 49, 3–19. <http://doi.org/10.3414/ME0630>
- Kobak, K. A., Stone, W. L., Wallace, E., Warren, Z., Swanson, A., & Robson, K. (2011). A Web-Based Tutorial for Parents of Young Children with Autism: Results from a Pilot Study. *Telemedicine and E-Health*, 17, 804–808. <http://doi.org/10.1089/tmj.2011.0060>
- Korus, M., Cruchley, E., Stinson, J. N., Gold, A., & Anthony, S. J. (2015). Usability testing of the Internet program: “Teens Taking Charge: Managing My Transplant Online.” *Pediatric Transplantation*, 19, 107–117. <http://doi.org/10.1111/petr.12396>
- Kushniruk, A. (2002). Evaluation in the design of health information systems: application of approaches emerging from usability engineering. *Computers in Biology and Medicine*, 32, 141–149. <http://doi.org/S0010482502000112> [pii]
- Kushniruk, A. W., & Patel, V. L. (2004). Cognitive and usability engineering methods for the evaluation of clinical information systems. *Journal of Biomedical Informatics*, 37, 56–76. <http://doi.org/10.1016/j.jbi.2004.01.003>
- Lee M, Morfini M, Schulman S, I. J. (2001). Factor VIII and Factor IX of the Scientific and Standardization Committee of the International Society on Thrombosis and Haemostasis, 1–9. Retrieved from http://c.ymcdn.com/sites/www.isth.org/resource/group/d4a6f49a-f4ec-450f-9e0f-7be9f0c2ab2e/official_communications/fviiiipharco.pdf
- Manco-Johnson, M. J., Abshire, T. C., Shapiro, A. D., Riske, B., Hacker, M. R., Kilcoyne, R., ... Evatt, B. L. (2007). Prophylaxis versus episodic treatment to prevent joint disease in boys with severe hemophilia. *The New England Journal of Medicine*, 357, 535–544. <http://doi.org/10.1056/NEJMoa067659>
- Monkman, H., Griffith, J., & Kushniruk, A. W. (2015). Evidence-based Heuristics for Evaluating Demands on eHealth Literacy and Usability in a Mobile Consumer Health Application (Vol. 216, pp. 358–362). <http://doi.org/10.3233/978-1-61499-564-7-358>
- Nielsen, J. (1993). *Usability Engineering*. San Diego, CA: Academic Press.

- Nielsen, J. (2012). Usability 101: Introduction to Usability. Retrieved May 30, 2015, from <http://www.nngroup.com/articles/usability-101-introduction-to-usability/>
- Nielsen, J., & Landauer, T. K. (1993). A mathematical model of the finding of usability problems. In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '93* (pp. 206–213). <http://doi.org/10.1145/169059.169166>
- Nilsson, I. M., Hedner, U., & Ahlberg, A. (1976). Haemophilia prophylaxis in Sweden. *Acta Paediatrica Scandinavica*, *65*, 129–135. <http://doi.org/http://dx.doi.org/10.1111/j.1651-2227.1976.tb16525.x>
- Nolan, T., Dack, C., Pal, K., Ross, J., Stevenson, F. a., Peacock, R., ... Murray, E. (2015). Patient reactions to a web-based cardiovascular risk calculator in type 2 diabetes: a qualitative study in primary care. *British Journal of General Practice*, *65*(632), e152–e160. <http://doi.org/10.3399/bjgp15X683953>
- Peute, L. W. P., Spithoven, R., Bakker, P. J. M., & Jaspers, M. W. M. (2008). Usability studies on interactive health information systems; where do we stand? *Studies in Health Technology and Informatics*, *136*, 327–332. <http://doi.org/10.3233/978-1-58603-864-9-327>
- 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. *Journal of the American Medical Informatics Association*, *1*, 1–5. <http://doi.org/10.1093/jamia/ocv050>
- Rubin, J., & Chisnell, D. (2008). *Handbook of Usability Testing How to Plan, Design, and Conduct Effective Tests* (2nd ed.). Indianapolis, Indiana: Wiley Publishing, Inc.
- Running a Usability Test. (2014, May 15). Retrieved May 27, 2015, from <http://www.usability.gov/how-to-and-tools/methods/running-usability-tests.html>
- Sandelowski, M. (2000). Focus on Research Methods Whatever Happened to Qualitative Description? *Research in Nursing & Health*, *23*, 334–340. [http://doi.org/10.1002/1098-240X\(200008\)23:4<334::AID-NUR9>3.0.CO;2-G](http://doi.org/10.1002/1098-240X(200008)23:4<334::AID-NUR9>3.0.CO;2-G)
- Scandurra, I., Hägglund, M., Persson, A., & Ahlfeldt, R.-M. (2014). Disturbing or facilitating?--on the Usability of Swedish eHealth Systems 2013. *Studies in Health Technology and Informatics*, *205*, 221–225. <http://doi.org/10.3233/978-1-61499-432-9-221>
- Schade, A. (2013). Remote Usability Tests: Moderated and Unmoderated. Retrieved May 10, 2015, from <http://www.nngroup.com/articles/remote-usability-tests/>

- Soucie, J. M., Symons, J., Evatt, B., Brettler, D., Huszti, H., & Linden, J. (2001). Home-based factor infusion therapy and hospitalization for bleeding complications among males with haemophilia. *Haemophilia*, 7, 198–206. <http://doi.org/10.1046/j.1365-2516.2001.00484.x>
- Srivastava, A., Brewer, a. K., Mauser-Bunschoten, E. P., Key, N. S., Kitchen, S., Llinas, A., ... Street, A. (2013). Guidelines for the management of hemophilia. *Haemophilia*, 19, e1–e47. <http://doi.org/10.1111/j.1365-2516.2012.02909.x>
- Stinson, J., Gupta, a., Dupuis, F., Dick, B., Laverdiere, C., LeMay, S., ... Chan, C. Y. (2014). Usability Testing of an Online Self-Management Program for Adolescents With Cancer. *Journal of Pediatric Oncology Nursing*, 32(2), 70–82. <http://doi.org/10.1177/1043454214543021>
- Stinson, J., McGrath, P., Hodnett, E., Feldman, B., Duffy, C., Huber, A., ... White, M. (2010). Usability testing of an online self-management program for Adolescents with Juvenile Idiopathic Arthritis. *Journal of Medical Internet Research*, 12(3), 1–20. <http://doi.org/10.2196/jmir.1349>
- System Usability Scale (SUS). (2013, September 6). Retrieved May 27, 2015, from <http://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html>
- Teixeira, L., Ferreira, C., & Santos, B. S. (2007). Using task analysis to improve the requirements elicitation in health information system. *Conference Proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference, 2007*, 3669–3672. <http://doi.org/10.1109/IEMBS.2007.4353127>
- Teixeira, L., Ferreira, C., & Santos, B. S. (2012). User-centered requirements engineering in health information systems: A study in the hemophilia field. *Computer Methods and Programs in Biomedicine*, 106, 160–174. <http://doi.org/10.1016/j.cmpb.2010.10.007>
- Teixeira, L., Saavedra, V., Ferreira, C., & Sousa Santos, B. (2009). The user's role in the development process of a clinical information system: An example in hemophilia care. *Lecture Notes in Computer Science (including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 5619 LNCS, 922–931. http://doi.org/10.1007/978-3-642-02806-9_106
- Thompson, K., Rozanski, E., & Haake, A. (2004). Here, there, anywhere: remote usability testing that works. *CITC5 '04 Proceedings of the 5th Conference on Information Technology Education*, 132–137. <http://doi.org/10.1145/1029533.1029567>

- User Interface Design Basics. (2014, May 21). Retrieved May 28, 2015, from <http://www.usability.gov/what-and-why/user-interface-design.html>
- Valentino, L. A., Mamonov, V., Hellmann, A., Quon, D. V, Chybicka, A., Schroth, P., ... Wong, W.-Y. (2012). A randomized comparison of two prophylaxis regimens and a paired comparison of on-demand and prophylaxis treatments in hemophilia A management. *Journal of Thrombosis and Haemostasis : JTH*, *10*, 359–67. <http://doi.org/10.1111/j.1538-7836.2011.04611.x>
- Virzi, R. a. (1992). Refining the Test Phase of Usability Evaluation : How Many Subjects Is Enough ? *Human Error and Clinical Systems*, *34*, 457–468. <http://doi.org/10.1177/001872089203400407>
- Wakefield, B., Pham, K., & Scherubel, M. (2015). Usability Evaluation of a Web-Based Symptom Monitoring Application for Heart Failure. *Western Journal of Nursing Research*, *37*, 922–934. <http://doi.org/10.1177/0193945914568813>
- Wang, J.-D., Chan, W.-C., Fu, Y.-C., Tong, K.-M., Chang, S.-T., Hwang, W.-L., ... Tsan, Y.-T. (2015). Prevalence and risk factors of atherothrombotic events among 1054 hemophilia patients: A population-based analysis. *Thrombosis Research*, *135*, 502–507. <http://doi.org/10.1016/j.thromres.2014.12.027>
- What Is Hemophilia? (2013). Retrieved July 28, 2015, from <http://www.nhlbi.nih.gov/health/health-topics/topics/hemophilia>
- Wolpin, S. E., Halpenny, B., Whitman, G., McReynolds, J., Stewart, M., Lober, W. B., & Berry, D. L. (2014). Development and usability testing of a web-based cancer symptom and quality-of-life support intervention. *Health Informatics Journal*, *21*, 10–23. <http://doi.org/10.1177/1460458213495744>
- Yen, P.-Y., & Bakken, S. (2012). Review of health information technology usability study methodologies. *Journal of the American Medical Informatics Association*, *19*, 413–422. <http://doi.org/10.1136/amiajnl-2010-000020>
- Yen, P.-Y., Sousa, K. H., & Bakken, S. (2014). Examining construct and predictive validity of the Health-IT Usability Evaluation Scale: confirmatory factor analysis and structural equation modeling results. *Journal of the American Medical Informatics Association : JAMIA*, *21*, e241–e248. <http://doi.org/10.1136/amiajnl-2013-001811>

Appendix A: Task List

Participants will start going through **Task 1-3**.

Task 1: Patient data entry

You registered an account with the WAPPS-Hemo service and want to enter 2 patients' data and their **first 3** blood sample draws. Then you will add a subsequent blood sample draws data to **one patient**. Your objective is to create a patient list of minimum two patients with minimum one subsequent infusion.

Task 2: Complete a patient's blood sample draws

Patient #1 will receive a 4th blood sample draw. Add that information to the patient's profile. Keep in mind that you might want to add new blood sample data in the future.

Task 3: Change patient's data and view results

You noticed that for Patient #2 his last blood sample draw data is not correct. Correct the information stored to 0.031 IU/ml.

Also, you want to see the results of another patient.

***Note:** While going through the steps, please think out loud. The observer will want to know how you will navigate within the website, enter the data, and your thought process during the task.*

Appendix B: Background Questionnaire

1. What is your profession/role at the center?
 - a. Hematologist/Physician
 - b. RN/Nurse Practitioner
 - c. Administrator
 - d. Other (please specify):

2. What is your age?

3. Do you currently use any PK calculator?
 - a. Yes
 - b. No

4. Have you used a PK calculator in the past?
 - a. Yes
 - b. No

If Yes, for how long?

5. Do you have experience with the WAPPS-Hemo service?
 - a. Yes
 - b. No

If Yes, for how long?

Appendix C: Introduction Script

Script¹:

Hi, _____. My name is Islam, and I'm going to walk you through this session today. I have some information for you and I'm going to read it to make sure that I cover everything.

As you know, we're asking people to participate in the usability of the WAPPS-Hemo service to gain more knowledge about the flaws that might be in the system and to get ideas on how to improve and develop the service. The session will take about 45 minutes.

The purpose of the WAPPS is that it will allow the pharmacokinetic (PK) parameters of an individual to be estimated with a decreased number of blood samples.

Therefore, PK assessment will be easier. It will also improve the knowledge about the PKs of factors VIII and IX.

The first thing I want to make clear is that we're testing the service, not you. You can't do anything wrong. In fact, this probably the one place you don't have to worry about making mistakes.

I'm going to ask you as much as possible to try to think out loud. Try to say what you're looking at and what you are trying to do. This will be very beneficial to the research.

¹ *Rocket Surgery Made Easy* © 2010 Steve Krug

If you have any questions as we go along, just ask them. I may not be able to answer them right away, since we're interested in how people do when they don't have someone sitting next to them to help. But if you still have any questions when we're done, I'll try answer them.

With your permission we're going to record what happens on the screen and our conversation. The recording will be anonymous and only be used to help us figure out how to improve the service.

Appendix D: SUS Survey

		Strongly Disagre				Strongly Agree
1	I think that I would like to use this system frequently.	1	2	3	4	5
		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2	I found the system unnecessarily complex.	1	2	3	4	5
		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3	I thought the system was easy to use.	1	2	3	4	5
		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4	I think that I would need the support of a technical person to be able to use this system.	1	2	3	4	5
		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5	I found the various functions in this system were well integrated.	1	2	3	4	5
		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
6	I thought there was too much inconsistency in this system.	1	2	3	4	5
		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
7	I would imagine that most people would learn to use this system very quickly.	1	2	3	4	5
		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
8	I found the system very cumbersome to use.	1	2	3	4	5
		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
9	I felt very confident using the system.	1	2	3	4	5
		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

10	I needed to learn a lot of things before I could get going with this system.	1	2	3	4	5
		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Appendix E: Post-Testing Debriefing

Follow up on any particular concerns or problems the participant had.

- Did you have any problems completing the tasks?
- Did you have any problems navigating through the system?
- What part of the service would you want to see different?
- Anything to be improved?

Appendix F: Summary of Results

SUS Scores of Iterations 1 & 2

SUS Item	Iteration 1	Iteration 2
	Subjects (n=5)	Subjects (n=5)
	M (SD) ^a	M (SD) ^a
1. I think that I would like to use this system frequently.	3 (0.71)	3.0 (0.71)
2. I found the system unnecessarily complex.	2.2 (1.64)	3.2 (1.30)
3. I thought the system was easy to use.	2.6 (1.14)	3.0 (1.22)
4. I think that I would need the support of a technical person to be able to use this system.	3.4 (0.55)	3.0 (1.58)
5. I found the various functions in this system were well integrated.	2.6 (0.55)	2.8 (1.30)
6. I thought there was too much inconsistency in this system.	3.0 (1.22)	3.2 (1.30)
7. I would imagine that most people would learn to use this system very quickly.	3.2 (0.84)	2.8 (1.30)
8. I found the system very cumbersome to use.	2.6 (1.52)	3.2 (1.10)
9. I felt very confident using the system.	3.0 (1.0)	2.6 (1.14)
10. I needed to learn a lot of things before I could get going with this system.	2.6 (1.67)	2.4 (1.52)
^a 1 = strongly disagree, 5=strongly agree		

SUS Sub-scale Scores			
		Iteration 1	Iteration 2
Sub-scale	SUS Item	Mean (SD)	Mean (SD)
Usability	2,3,5,6,7,8,9,10	2.73 (9.58)	2.92 (9.78)
Learnability	1,4	3.2 (1.25)	2.70 (1.98)

Appendix G: NVivo 10 Coding for Iteration 1

The screenshot shows the NVivo 10 software interface. The 'Nodes' pane on the left lists 'Nodes', 'Relationships', and 'Node Matrices'. The main window displays a table of nodes with the following data:

Name	Sources	References	Created On	Created By	Modified On
Interface Problems	5	88	30/03/2015 6:43 AM	I	01/04/2015 9:53 AM
Negative Comments	3	5	30/03/2015 12:19 PM	I	01/04/2015 9:23 AM
Neutral Comments	5	12	30/03/2015 12:19 PM	I	01/04/2015 9:43 AM
Positive Comments	4	12	30/03/2015 12:19 PM	I	01/04/2015 9:44 AM
Suggestions for Improvements	5	17	30/03/2015 6:44 AM	I	01/04/2015 9:44 AM
Technical Problems	4	5	30/03/2015 6:44 AM	I	30/03/2015 8:25 PM

Appendix H: NVivo 10 Coding for Iteration 2

The screenshot shows the NVivo 10 software interface. The 'Nodes' pane on the left lists 'Nodes', 'Relationships', and 'Node Matrices'. The main window displays a table of nodes with the following data:

Name	Sources	References	Created On	Created By	Modified On
Interface Problems	5	14	13/05/2015 8:29 PM	IB	15/05/2015 10:01 PM
Negative Comments	2	4	13/05/2015 8:30 PM	IB	15/05/2015 9:46 PM
Neutral Comments	4	11	13/05/2015 8:30 PM	IB	15/05/2015 10:00 PM
Positive Comments	5	18	13/05/2015 8:30 PM	IB	15/05/2015 9:59 PM
Suggestions for Improvements	7	15	13/05/2015 8:30 PM	IB	15/05/2015 10:01 PM
Technical Problems	3	3	13/05/2015 8:31 PM	IB	15/05/2015 10:01 PM