A Practical Framework for Usability Evaluation of an IT-System in Use

- *Developed for a Pharmaceutical Company*

John Lunde Flennmark & John Ström

Master's Thesis

Department of Design Sciences
Lund University
ISRN:LUTMDN/TMAT-5100-SE

EAT 2007

This master thesis was financed by AstraZeneca R&D Lund.

Keywords
AstraZeneca, benefit market, business case, cost-benefit analysis, framework, guidelines for evaluation, severity rating, usability, usability engineering, usability evaluation, usability problems

Sökord (Swedish)
AstraZeneca, användbarhet, användbarhetsproblem, användbarhetsprocessen, användbarhetsutvärdering, business case, kostnadsnyttoanalys, marknadsvärde, ramverk, riktlinjer för utvärdering, severity rating

Abstract
This master thesis was written at the department of design sciences at Lund University.

In the thesis a framework for usability evaluation was developed. The framework is intended to aid the usability process at AstraZeneca's R&D department in Lund, and help the company to improve the usability of their IT-systems. It is also meant to be used to evaluate and find usability problems in a system in use. The framework is supposed to guide the staff at the R&D department through a cost-benefit analysis of the usability problems found in the evaluation, which is a part of the framework. The cost-benefit analysis helps form the basis for further development in a business case, where the benefit for fixing the problems can be calculated. The thesis was written in close cooperation with AstraZeneca staff and was written at AstraZeneca in Lund.

To insure that the framework has validity and reliability two case studies were performed at the company. The case studies helped the developers to extend and improve the framework.

At the end of the thesis, a final framework proposal is presented where guidelines for using the framework are provided.

Glossary
AZ
Short for AstraZeneca

Benefit market
The maximum benefit that can be generated by a change.

Business case
A structured proposal for a business change. A typical prerequisite for the initiation for a large project.

Case study
A way to test a theory in practice.

CBA
Short for cost-benefit analysis. A calculation of initial and ongoing expenses versus expected return.

Clinical study
A study where a medication is tested.

Evaluator
Person with usability skills that can conduct a usability evaluation.

Framework
A set of guidelines to aid in a process.

FTE
Short for full time equivalent. A measure for how much one full time employee would cost.

Investigator
An m.d. that is involved in a clinical study.

M.d.
Short for medical doctor.

Monitor
Person that works with clinical studies. Is employed to check that the investigator performs the clinical study correctly. Performs SDV checks.

R&D
Research and Development, a department within AstraZeneca.

SDV
Short for source data verification. A check that controls that the information supplied by the investigator is correct.

StudyAdmin
A fictive name for one of the systems in the case studies.

StudyCapture
A fictive name for one of the systems in the case studies.

Super-user
A person that has a supporting role in the company. The support given is intended to help the monitors with their work. The super-user has extensive knowledge in their specific program.

SUS
Short for system usability scale. A scale that provides a result of the attitude towards the usability of a system.

System administrator
A person that knows the system well and have administrator rights within it.

System owner
The person that is responsible for the system and has an important vote in managerial decisions.

Executive summary

The usability of an IT-system has a strong impact on the quality of the work that is performed with the system. Although, it can often be hard to evaluate the usability of a system since it is hard to measure. A system with poor usability can for example, result in that the users feel that the system is hard to work with and that more mistakes are made and have to be corrected. The consequences of poor usability often result in a loss of time. And as we all know, time is money.

The main question of this thesis is; *is it possible to devise a general framework that guides AstraZenecas R&D staff in the usability evaluation process*? A framework can support a standardization of the usability work in an organization. If the usability engineering process is standardized in an organization and improving the usability becomes an iterative process, money can be saved and the user satisfaction can be increased. A practical interpretation of the main question is, that a framework for usability evaluations of IT-systems that already are in use in the organization, should be developed. The evaluation should also include a cost-benefit analysis.

The goal with this master thesis was to answer the main question and to develop the framework, if possible. It was seen as vital that the framework fit into the organization at AstraZeneca and that it would be usable for the staff at the company. The master thesis builds on a prior thesis that was conducted at the Research and Development (R&D) department at AstraZeneca in Lund, in the spring of 2005.

To investigate if it was possible to devise a framework that fit into the organization at AstraZeneca, a first framework proposal was developed early on in the work process. The first framework proposal was divided into four main phases according to the following:

- Background studies
- Usability Evaluation
- Usability rating
- Cost-benefit analysis

The four phases of the first framework were also used in the final framework proposal. An application of the framework results in a number of specified and prioritized usability problems. Additional results are, a cost-benefit analysis of each of the most critical usability problems and the results from an attitude survey concerning the usability of the system.

A requirement from AstraZeneca was that the framework should be executable by one evaluator within 40 man-hours. An additional goal was that the framework results should provide the basis for a business cases were the actual benefit for fixing the usability problems could be further analyzed.

The first framework proposal was tested in two case studies were the evaluators tried to fulfill all the goals except for the timeframe for the evaluation. The timeframe was eliminated to give the evaluators time to further develop the framework details during the case studies at the same time as they were testing the first framework proposal. The case studies were performed on two IT-systems that are used in AstraZenecas clinical studies.

A requirement for staff members to be able to use the framework is that they have experience in the usability area. There is no need to understand the AstraZeneca organization, but it helps. If the user of the framework does not have to spend time getting to know the organization, more time can be spent on the other activities within

the framework. The experience in the usability area that the user possesses plays an extra large part in the actual usability evaluation phase. This thesis provides some guidelines on how to proceed with an evaluation, but in the end it is up to the user how to perform the usability inspections. The framework is constructed in such a way that the user can conduct the evaluation as he/she pleases as long as it results in a number of usability problems that can be further evaluated through cost-benefit analysis.

The case studies showed that the framework was practically usable and that it would fit into the organization. Changes were made to the framework based on the case studies and the final framework proposal is presented as a result. Its users should still improve the framework iteratively so that it adapts to the needs of the organization but the main goal of the thesis can be considered to have been fulfilled. The question *Is it possible to devise a general framework that guides AstraZenecas R&D staff in the usability evaluation process*? was not completely answered though. Since its developers tested the framework, no conclusions can be drawn on if it is going to work with the real users. This fact still stands even though the theories that lay the groundwork for the framework are well thought through and tested. The modification and adaptation of the theories may have created problems that have not been evaluated in this thesis.

The result of this master thesis is a framework that is practically applicable and that provides results that have a high enough validity and reliability to be considered credible and usable. The framework supports usability evaluation of IT-systems in-use and provides guidelines for prioritizing usability problem. Additionally it supports the process of conducting a cost-benefit analysis for the usability evaluation-process.

Sammanfattning (Swedish)

Användbarheten hos ett IT-system påverkar kvalitén på arbetet som utförs med det till en stor grad. Det är dock ofta väldigt svårt att utvärdera användbarheten för ett system. Dålig användbarhet hos ett IT-system innebär t.ex. att användaren tycker systemet känns svårjobbat eller att användaren lätt gör misstag som senare måste rättas. Alla sådana här faktorer kan översättas i förlorad tid. Och som vi vet, tid är pengar.

Examensarbetets huvudfråga är; *Är det möjligt att ta fram ett ramverk som kan stödja AstraZenecas R&D personal i användbarhetsutvärderingsprocessen?* Ett ramverk kan hjälpa till att standardisera användbarhetskonceptet i en organisation. Om användbarhetsarbetet standardiseras och blir till en iterativ process kan man både spara pengar och öka användarnas arbetstillfredställelse. I praktiken innebär frågan att det ramverk som ska utvecklas, ska kunna användas till att utvärdera användbarheten hos ett IT-system, som finns i organisationen sedan tidigare. Utvärderingen ska även resultera i en kostnads-nyttoanalys.

Målet med examensarbetet var att svara på huvudfrågan, samt att utveckla ramverket, om det ansågs möjligt. Viktigt var att ramverket skulle passa in i AstraZenecas organisation och att det skulle kunna användas av personalen på företaget. Examensarbetet bygger vidare på ett tidigare arbete som utfördes på R&D avdelningen på AstraZeneca i Lund, våren 2005.

För att testa tesen att det skulle gå att utveckla ett ramverk som passade AstraZeneca, gjordes en första ansats till ramverket tidigt i processen. Ramverket delades upp i fyra faser enligt följande:

- Bakgrundsstudier
- Användbarhetsutvärdering
- Användbarhetsgradering
- Kostnads-nyttoanalys

De fyra faserna ovan blev även använda i det slutgiltiga ramverksförslaget. En genomgång av faserna i ramverket resulterar i ett antal specificerade och prioriterade användbarhetsproblem. Det kommer även att resultera i en kostnads-nyttoanalys av de mest kritiska användbarhetsproblemen samt resultaten av en attitydundersökning kring användbarheten av systemet.

De fyra stegen som användes både i första ramverks förslaget och det slutgiltiga var:
Ett krav från AstraZeneca var att ramverket skulle kunna utföras på 40 mantimmar och av en utvärderare. Målet var även att det skulle resultera i underlag för vidare utveckling i ett business case där själva nyttan av att rätta upp felen som hittades kan analyseras vidare.

Ansatsen testades genom två fallstudier enligt specifikationerna, med undantaget av att tidsramen togs bort. Detta gjordes för att för att utvecklarna skulle ha tid att utveckla ramverket vidare, samtidigt som de testade upplägget på ramverket. Fallstudierna gjordes på två IT-system som används inom AstraZenecas kliniska studier.

För att kunna använda sig av ramverket krävs det att man har erfarenhet inom användbarhetsområdet. Man behöver inte ha god insikt i AstraZenecas organisation, men det hjälper till. Om användaren av ramverket slipper lägga tid på att sätta sig in i organisationen kan mer tid läggas på de övriga stegen. Erfarenheten hos användaren spelar speciell stor roll när själva utvärderingsfasen tar vid. Examensarbetet erbjuder vissa riktlinjer och tips på hur man kan gå till väga när man utvärderar, men i slutändan är det ändå upp till användaren hur fasen ska gå till väga. Ramverket är anpassat så att användaren ska kunna utföra utvärderingen på nästan vilket sätt som helst, så länge den resulterar i användbarhetsproblem som går att vidareutvärdera ur ett kostnads-nyttoperspektiv.

Fallstudierna visade att ramverket var praktiskt användbart och skulle kunna passa in i organisationen. Det gjordes även ändringar i ramverket efter fallstudierna för att ta fram det slutgiltiga förslaget som presenteras som resultat. Ramverket måste givetvis vidareutvecklas iterativt för att anpassas helt efter företagets förutsättningar, men huvudmålet med examensarbetet är uppfyllt. Frågan *Är det möjligt att ta fram ett ramverk som kan stödja AstraZenecas R&D personal i användbarhetsutvärderings-processen?* blev inte helt besvarad dock. I och med att ramverket testades av utvecklarna själva, finns det inget underlag för att säga att det fungerar med riktiga användare. Detta faktum kvarstår även om de teorier som används som grund till examensarbetet är väl genomarbetade och väl testade. Viss modifiering och hopslagningen av olika teorier kan skapa problem som inte har blivit utredda i denna uppsats.

Resultatet av examensarbetet blev ett ramverk som är praktiskt användbart och genomförbart och där resultatet kommer att ha tillräckligt hög validitet och reliabilitet för att anses trovärdigt. Ramverket ger stöd för användbarhetsutvärdering av

befintliga IT-system samt ger riktlinjer för hur användbarhetsproblem kan prioriteras. Vidare ges stöd för att göra en kostnads-nyttoanalys av användbarhetsprocessen.

Acknowledgements

We would like to take the opportunity to thank our supervisor at AstraZeneca, Pia af Klercker, for helping us through the work with this master thesis and for making our time at AstraZeneca in Lund pleasant.

We would also like to show our appreciation to Veronica Sällberg, our second supervisor at AstraZeneca for helping us with many technical details, opinions and an enjoyable car drive to Mölndal.

Mikael Palmblad and Sara Lignell have helped us a lot with the actual usability work and the deriving of the model for the cost-benefit analysis. Knowing that you think that our work is valid and has a use within the organization means a lot to us.

We also thank our supervisor at Lund University, Joakim Ericsson, who has guided us through this master thesis and has helped us along the way.

Mattias, Lisa and Patricia, our dining partners at AstraZeneca, you have livened up the atmosphere and have made the darkest of times lighter.

We would like to thank everyone else within AstraZeneca that have helped us during this master thesis. All your input has helped us a great deal.

Last but not least, we would like to thank our families, friends and girlfriends for all the support that you have given us through our studies at Lund University. We know that we have not always been the nicest of guys during stressful times, but you have all hung in there knowing what great guys we really are beneath our rough exterior.

Table of contents

1 Introduction

1.1 Background

Any major company acting on the global market today is extremely dependent on IT-systems to support their business. As the competition between companies grows, the demands on the IT-systems increase as well. Today it is not enough that a system has the required functionality, it has to be effective and efficient to be competitive. One factor that contributes strongly to the quality of an IT-system with user-interaction is the usability of the system. Better usability can help reduce the costs associated with the system and increase it's effectiveness and efficiency.

As a major international pharmaceutical business AstraZeneca has high demands for IT-systems of high quality. There are high demands on pharmaceutical companies to have thoroughly tested the drugs that they supply the market with. The process for testing and developing drugs therefore has to be extremely controlled and well documented. Today most of the documentation and administration is done through, or with the support of IT-systems.

AstraZeneca has realized the value of increasing the usability of their IT-systems and are therefore continuously working on improving their usability engineering processes. One way to improve the usability engineering efforts and increase the awareness of usability at a company is to standardize the process. Based on the facts above this thesis was initiated with the aim to develop a framework for usability evaluation that is to be used by AstraZenecas Research and Development (R&D) staff in Lund. It was also seen as important that the usability evaluation was connected to a cost-benefit analysis to motivate the usability effort towards the rest of the organization.

1.2 Objective

The main objective of this master thesis is to investigate how the usability of IT systems at AstraZeneca (AZ) can be improved through providing guidelines for the usability evaluation process. The investigation is conducted through the development of a framework for usability evaluation that is to be used at the R&D department at AZ, Lund. It is also important that the framework is general and not only applicable at this department at AZ. For the framework to be practical and usable at the R&D department the following requirements on the framework need to be fulfilled:

- Practically usable for someone with experience in the usability area
- Support the process of finding usability problems
- Support a cost-benefit analysis of the usability evaluation
- Fit a timeframe of around 40 man-hours
- Fit the organization at AZ
- General and not only suit AZ's needs

In order of developing a framework that is generally usable, the development process needs to be scientific. There is also a need for evaluating the framework through case studies at AZ to evaluate how usable it is for the company.

Secondary objectives are to investigate if the framework can be used for:

- Evaluation of systems that have not been introduced at the company
- Comparing the usability of different systems with the same purpose

1.3 Defining the problem

Based on the objective of this thesis a few main problems were defined for investigation. The problems were defined so that investigating them would lead to a scientific process where the present research in the area would be studied and a solution would be developed and tested through case studies. The following questions define the problems that this thesis aims to investigate:

- The main question is; *is it possible to devise a general framework that guides AstraZenecas R&D staff in the usability evaluation process*? This question can be divided into a couple of smaller problems.
  1. Are there any general theories about usability and CBA that can be used in the development of the framework for usability evaluation?
  2. Is it possible to devise a framework that serves as a guide to usability evaluation and CBA at AZ?
  3. Can the framework guide the AZ staff in the usability evaluation process?

1.4 Target group

The target group for this thesis is people working with usability in the business world as well as the academic sphere. However the primary target group is AstraZeneca staff working with usability or IT-systems in general.

1.5 Limitations and problems

Early on in the process a decision was made, that if the timeframe was to be held, the secondary objectives could not be fulfilled. The decision was based on a conclusion that evaluating new systems and comparing systems with similar functionality would require different approaches than evaluating a system in use. A consequence of this was that the secondary goals could not be incorporated in the same framework and they were therefore only suggested as future research.

1.6 Usability at AZ

As a major international pharmaceutical business AstraZeneca has high demands for well functioning IT-systems to support and enable the organization to work more efficiently. In order for the IT-systems to provide the best possible support for the company, a usability effort is needed.

1.6.1 General usability work

At the time that this thesis was written, there was a small but dedicated group working with improving the usability of IT-systems at AstraZeneca's R&D department. The staff members with usability experience were not concentrated to a separate unity; they were spread-out throughout the organization. Most of the people that were working with usability were not solely dedicated to that function either; they had other responsibilities in the organization as well.
In order to keep up to date with usability within the organization the staff had formed a usability network to exchange experience, knowledge and to plan future efforts. The staff had extensive knowledge in working with usability but the standardized work procedure could be further developed and improved.

1.6.2 The prior framework

Prior to this thesis, another student had investigated usability at AstraZeneca R&D. Her master thesis “Usability Evaluation of a System in Use” (Vagelin, 2005) resulted in a plan for evaluation of system usability. At the time that the work on this thesis was started AstraZeneca’s staff were still referring to Vagelin’s thesis. The staff at AstraZeneca R&D, who initiated this thesis, expressed a wish for it to build on Vagelin’s work. A summary of Vagelin’s *Plan for Evaluation of System Usability* can be found in chapter 3.8.

2 Method and structure

This chapter gives the reader a chance to understand how the work in this thesis has been performed and how the thesis is structured.

2.1 Work process

This thesis was performed in an orderly and a scientific manner.

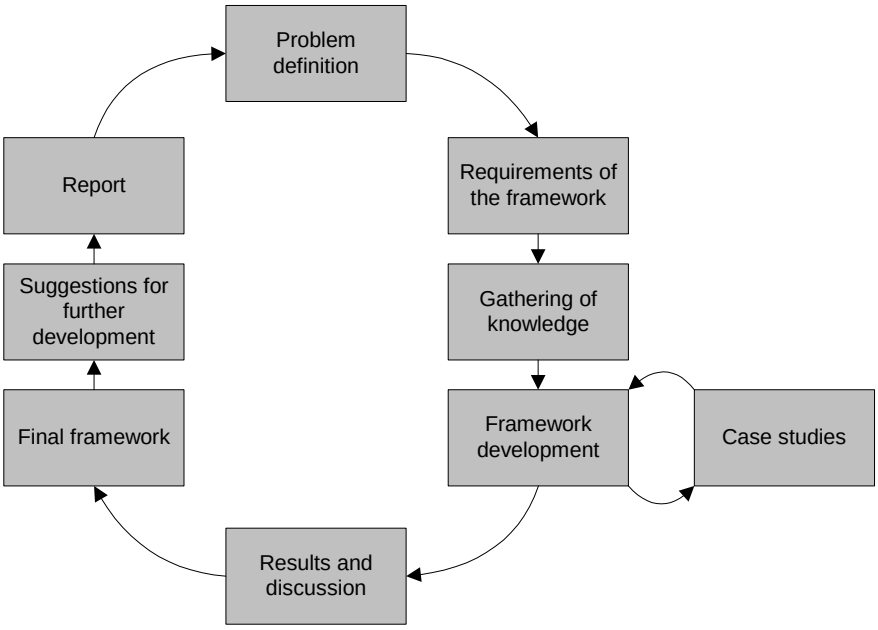

Figure 2.1 – The work process for this master thesis

Early focus was on developing a first framework proposal. This was done to have something to build on during the project. During this first development an extensive literature study was started. New research and research by some of the most prominent authors within the usability field was studied and documented in this thesis.

During the literature study the framework was reworked and enhanced to more fit its purpose within AZ. Choices, made to enhance the framework, had to be motivated through the literature study and decisions had to be carefully motivated. Usability-competent staff within AZ and at the university was consulted in order to validate the usefulness of the framework.

When the framework was well enough motivated and worked through two case studies were prepared. Two suitable systems were chosen and it was made clear that the purpose was to test the framework, not the systems. The systems were only meant to act as tools for evaluating the framework. StudyAdmin and StudyCapture were the fictive names of the two systems.

The case studies were performed according to the framework proposal with one exception. The framework was to be evaluated so extra time was given in order for the evaluators to evaluate the steps in the framework itself. All progress and all comments were documented and were, later on, the subject for discussion.
Case studies were chosen as the testing method because it is a good way of testing the theory in a practical way. The case studies give the evaluators a chance to test

the first framework proposal. It also gives them a chance to evaluate how long each step of the procedure is supposed to take.

A background study of each system was performed where several interviews were held and a demonstration was attended. Attitude surveys were sent out and later analyzed. The evaluator conducted a heuristic evaluation, where usability problems were found. These problems were later on the subject for a severity rating process, where several stakeholders served as input to the rating. The five problems with the highest rating then moved on to be analyzed in a cost-benefit aspect. The cost-benefit analysis involved the evaluator and suitable AZ staff. The result of the case study was presented in a document and a presentation was held.

When the two case studies had been conducted a discussion was held where enhancement suggestions were discussed and where all choices made during the development phase were motivated. The discussion was important in order to really validate the results, and to make sure that they were usable as well as fully compatible with the requirements stated by AZ before the project started.

When the discussion was finished the final framework was presented in a separate chapter along with explanations on how to perform it. The chapter was written so that it could be cut out of the thesis. This decision was made so that it would be easier to find and access for anyone interested in using the final framework.

The project ended with a chapter dealing with enhancement suggestions and other problems that arose during the project. A couple of presentations of the work in this thesis were held at AZ and one was held at the university.

2.2 Thesis structure

In order to help the reader, an overview of the thesis is now presented.

Chapter 1: Introduction
The background for the thesis along with its goals and objectives is presented in this chapter. The work with usability within AstraZeneca is also described.

Chapter 2: Method and structure
Contains a description of the structure of the thesis and a short description of each chapter is given.

Chapter 3: Theory
This chapter presents all the theory that this thesis is based on. This chapter is written so that the reader gets a broader field of knowledge before getting in to the analysis part of the thesis.

Chapter 4: First framework proposal
This chapter gives the reader a first insight to what the framework might look like in the end. The framework presented here is the one the two case studies, later on, are based on. Motivations are given to all choices made and explanations are given to all method descriptions.

Chapter 5: Case study – StudyCapture
The first of the two case studies is described in this chapter. The results along with a brief case discussion are presented. The system called StudyCapture is the focus of the case study.

Chapter 6: Case study – StudyAdmin
The second of the two case studies is described in this chapter. The results and a brief case discussion are presented. The system called StudyAdmin is the focus of the case study.

Chapter 7: Framework discussion
In this chapter the limitations and decisions for the final framework proposal are discussed. Every choice made is motivated and explained. Questions that have been raised during the project are answered.

Chapter 8: Results – The final framework
The final framework is presented along with explanations on how to execute the framework. This chapter is written so that it can be cut out of the rest of this paper. It is written this way to make the framework easier to use.

Chapter 9: Conclusions
Other problems that have arisen during this project are discussed. Enhancement suggestions and further development is also briefly discussed.

Chapter 10: References
A list of the references that were used in this thesis.

3 Theory

3.1 Usability

Usability is an abstract concept that can be hard to define and understand. Today there are several different definitions of the concept; one of these is presented below. To understand the concept of usability it is important to know which factors to consider in a human performance situation. According to Rubin (1994) the Bailey's Human performance model describes the three major components to consider in a human performance situation, the factors are:

1. The human
2. The context
3. The activity

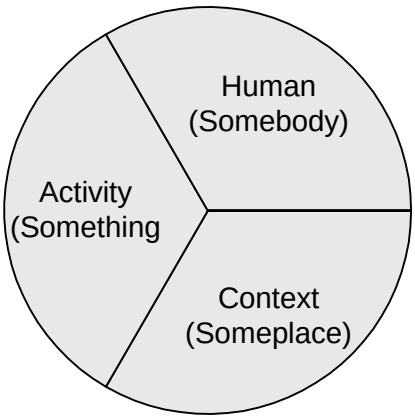

Figure 3.1 - Bailey's Human Performance Model. Authors' interpretation from Rubin (1994)

Rubin states that since the development of a system or product is an attempt to improve human performance in some area it is important for designers to consider these three components during the design process. For a system or product to be successful it is important that all three factors are considered, however this is not always the case. Traditionally the emphasis has been on the activity component and much less on the human and the context components.

Rubin further argues that one way of achieving better usability in a product or system is through user-centered design. The term user centered design is used to describe an approach that has been around for decades under different names, such as human factors engineering, ergonomics and usability engineering. He further states that there are many different approaches to user-centered design that can be used to achieve better usability. According to Rubin, the authors Gould and Lewis presented three key characteristics of a user-centered design. These three principles are:

1. An early focus on users and their tasks.
2. Empirical measurement of product usage
3. Iterative design whereby a product is designed, modified, and tested repeatedly.

3.1.1 Usability definition

The ISO definition of usability is one of many definitions of the concept, however it is one of the most accepted and used definitions today. In this thesis the ISO definition will be used whenever usability is discussed. This definition is also used in Anna Vagelin's master thesis *Usability evaluation of a system in use,* and since this thesis builds on the work of Vagelin we have adopted the same definition so that the work will be compatible.

ISO definition of usability:
Usability is the **effectiveness**, **efficiency**, and **satisfaction** with which specified users achieve specified goals in particular environments
Effectiveness - the accuracy and completeness with which specified users achieve specified goals in particular environments.
Efficiency - the resources expended in relation to the accuracy and completeness of goals achieved.
Satisfaction - the work comfort and acceptability of the work system to its users and other people affected by its use.

Table 3.1 - ISO definition of usability (ISO, 1998)

3.1.2 Motivations for usability

The motivation to create interfaces with better usability arises from the growing recognition of how an interface design can benefit the user greatly. However the motivations can differ widely depending on who the users and stakeholders are in the project. For example different system types yield different motivators for better usability.

Examples of system types:

- Life critical systems
- Industrial and commercial uses
- Office, home and entertainment applications

The focus in this thesis will be on systems for industrial and commercial use since the framework is developed for a commercial company. The motivation for usability in these types of systems is that costs shape judgments.

Better interface usability can improve the following system factors:

- Training time and ease of learning
- Speed of performance
- Error rates (& corrections) – system lifetime cost
- Handling high volumes of transactions
- Operator fatigue, stress, burnout
- Reduction in cost

Better usability can also benefit the company that is developing the application through increasing sales and customer satisfactions as well as helping them avoid building unnecessary functions.

3.1.3 Usability measures

For a usability effort to be effective there has to be some way to measure it's success. According to Gould and Lewis (Rubin, 1994) empirical measurement of product usage is one of the pillars of usability.
The ISO 9241 standard is not very easy to interpret into measurements that are practically useful. Therefore it is important to translate it into something more useful and more suitable for practical evaluation. Ben Shneiderman (2004) has defined a new set of usability measures in his book *Designing the user interface*. He came up with five more suitable categories for practical evaluation:

- Time to learn
- Speed of performance
- Rate of errors by users
- Retention over time
- Subjective satisfaction

These criteria's are much easier to directly translate into specific usability goals by asking more specific questions like; how long does it take to carry out the benchmark tasks? The ISO 9241 also deals with effectiveness, something that Shneiderman does not include in these measures.

3.1.4 Usability engineering lifecycle models

There are many different lifecycle models that are used in the software development process. They are used to determine and plan which activities that are to be performed and when to perform them. The models that are used in the usability engineering field today have evolved from traditional software development processes such as *The waterfall lifecycle model* and *The spiral lifecycle model* that are described in detail by Preece, Rogers and Sharp (2002). Increasing focus on user needs and usability has stimulated the development of models to fit a more user centered development process. Two alternative models that are used in the human-computer interaction field today according to Preece et al. (2002) are described below.

The star lifecycle model
Hartson and Hix proposed the Star lifecycle model in 1989; the model was based on empirical studies of how interface designers planned their work. The Star lifecycle unlike other lifecycle models does not specify the order that activities should be conducted; instead the user of the model is free to plan their work according to their preferences as long as every activity is followed by an evaluation.

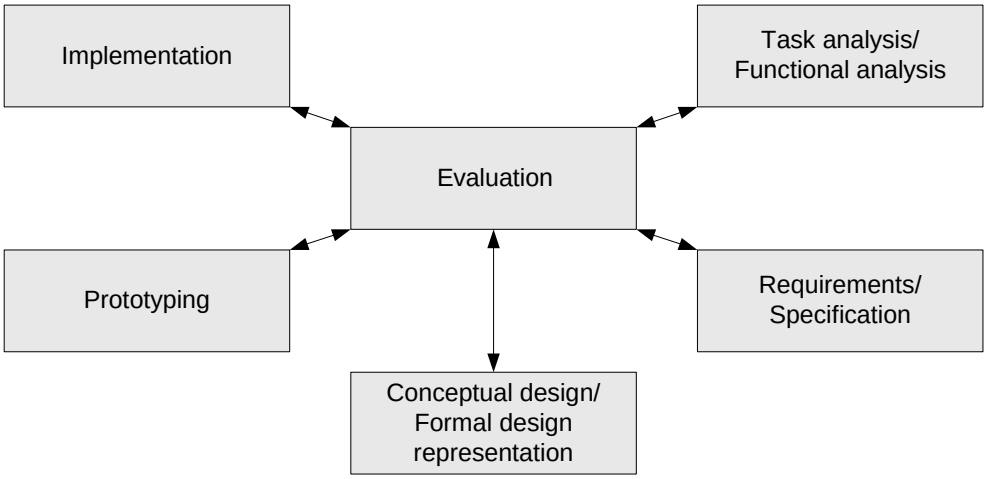

Figure 3.2 – The star lifecycle model. Authors' interpretation from Preece et al. (2002).

ISO 13407 - Human centered design processes for interactive systems

The ISO standard 13407 specifies a model for human centered design; it "provides guidance on achieving quality in use by incorporating user centered design activities throughout the life cycle of interactive computer-based systems" (UsabilityNet, 2006). It defines four key activities that must be performed at the early stages of a project:

- Understand and specify the context of use
- Specify the user and organizational requirements
- Produce design solutions
- Evaluate designs against requirements

The model also suggests an iterative approach to the design process.

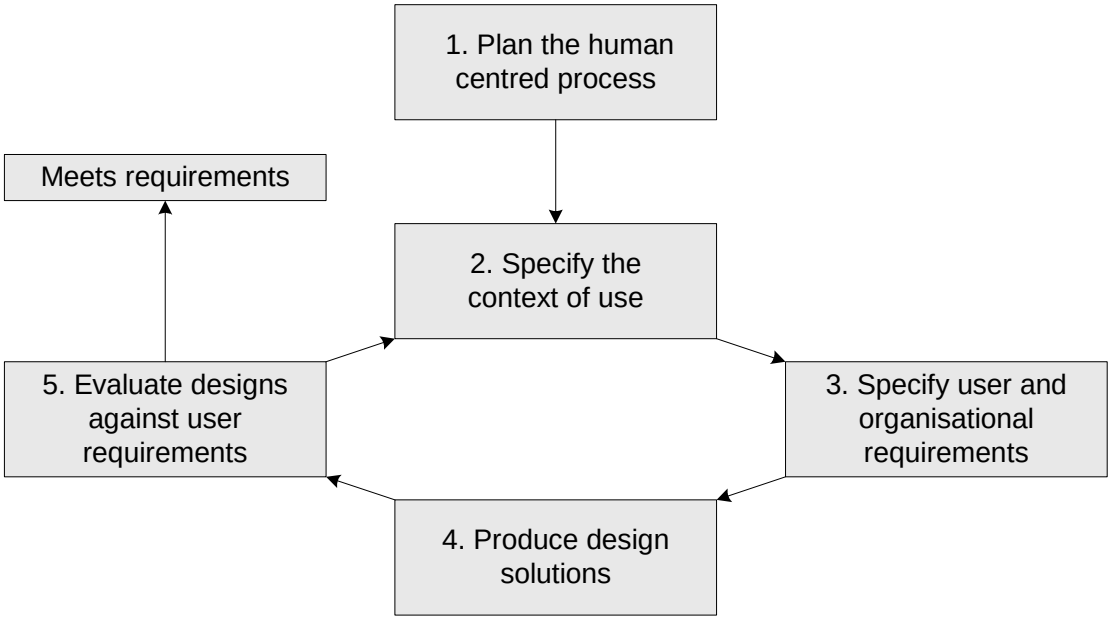

Figure 3.3 - The interdependence of user centred design activities. Authors' interpretation from UsabilityNet (2006)

3.2 Institutionalization of usability

When a usability program is being introduced into an organization, or if the way usability is practiced within an organization is altered, it is important that the changes fit the organizations needs and framework. Rubin (1994) stresses that if a usability-program is going to be effective and have high quality it is of high importance that is supported by everyone in the organization and that it is not enough that it is mandated by experts and leaders. Rubin goes on to write that it is better to allow a usability program to mature slowly and gradually improve within an organization than to make an effort to introduce a strategy to immediately attack all usability problems. According to Rubin a good strategy for introducing a usability-program is to devise an initial framework that is used and improved iteratively over time.

In his book *Institutionalization of usability*, Eric Schaffer (2004) argues that when an organization selects a user-centered methodology to fit their software development lifecycle, it is important that the methodology meets the following criteria:

- *It must be comprehensive.* The process cannot rely on usability testing solely; the whole software lifecycle must be addressed.
- *It must be user-centered.* This means that it must be firmly based on user needs and designing for optimal user experience and performance. Additionally it must also actually access representative users for data to support the process.
- *It must have a complete set of activities defined and deliverable documents required.* It should not only be based on a loose collection of ideas but a specific set of activities and documentation that supports the process.
- *It must fit corporate realities.* The user-centered design process must include steps that bring together views and ideas of the stakeholders within the organization. There must be activities that ensure that all key stakeholders get to contribute and fell heard.
- *It must be a good fit for the organization's size and criticality of work.* Large organizations that develop large and critical applications need a more thorough process and more detailed documentation.
- *It should be supported.* The implementation of a process requires training, templates, tools, standards and support services.
- *It must be able to work with the current development lifecycle.* The process has to fit the current development lifecycle. For example communication, coordination and handoffs between the different types of development staff must work properly.
- *It should ensure that the methodology has a cross-cultural localization process* where the design is evaluated for languages and cultural issues (if the organization has cross-cultural or international development)

3.3 Guiding principles

Research in the usability field has resulted in several sets of guidelines, heuristics and factors that should be considered when designing a system with good usability. These principles should be considered a rule of thumb or best practice and are in no way a shortcut to good usability in interfaces. To ensure better usability these principles have to be combined with experience and knowledge in the usability field. The principles are specified very generally and have to be adapted to fit the type of system that is being designed, in this case software supporting the clinical studies process.

3.3.1 Cognitive interpretation – Design principles

The usability design principles are based on theory about human cognition as well as extensive experience in the usability field and common sense. The purpose of the principles is to provide guidance and support designers in the interaction design process. The most famous and most commonly used set of principles are those proposed by Donald Norman (1998). The principles are described briefly below:

Visibility
The principle of visibility stresses the importance to make the functions, which are important and relevant at the moment, visible and making the functions, that are not relevant at the time, not visible. Doing this reduces the effort for the users and makes it easier for them to decide on their next action. Visibility also relates to the positioning and relative placement of controls in the interface.

Feedback
The concept of feedback relates to the visibility concept. It is important that the users receive feedback on all their actions in an interface. This helps the users find out the status of the system as well as what they have accomplished and which actions they should proceed with.

Constraints
This concept illustrates the importance of placing constraints on the kind of user interaction that can take place at a given time to make the decision easier for the user. According to Norman there are three categories of constraints: physical, logical and cultural. Physical constraints refer to the way that physical objects restrict the actions that can be made by the user. Logical constraints rely on peoples understanding of how the world works. These constraints are achieved through making it easy for the user to understand what their next actions should be. Cultural constraint rely on conventions of the user, they have to be learned and can differ widely between user groups.

Mapping
Mapping refers to the relationship between controls and their effects in the world. The controls should have a natural placement in relation to each other as well as to the effects that they have. For example the arrow controls on a keyboard should always have the up-arrow on top, the down arrow at the bottom and so on.

Consistency
This refers to designing interfaces to have similar operations and use similar elements for achieving similar tasks. One of the benefits of making an interface consistent is that the user only needs to learn one type of action to perform a certain task, which will save a lot of time and effort.

Affordance
The term affordance is used to indicate whether the attributes of an object help the user to understand how to use it. For example, a button should invite pushing. Norman clarifies this concept by dividing affordance into two categories: perceived and real. Physical objects have real affordances, they can for example be grasped or pushed and their function is obvious and does not have to be learned. Screen based user interfaces on the other hand do not have real affordances; they have perceived affordances. Norman argues that it does not make sense to design user interfaces to have real affordances, they should instead be designed according to learned conventions.

3.3.2 Eight golden rules of interface design

The golden rules of interface design are a set of principles that have been proposed and refined over two decades by Ben Shneiderman. The principles serve as a guide to good interaction design but need to validated and tuned to fit a specific design domain such as the clinical studies support systems. (University of Texas, 2006)

1. Strive for consistency
Consistent sequences of actions should be required in similar situations; identical terminology should be used in prompts, menus, and help screens; and consistent commands should be employed throughout.

*2 Enable frequent users to use shortcuts***.**
As the frequency of use increases, so do the user's desires to reduce the number of interactions and to increase the pace of interaction. Abbreviations, function keys, hidden commands, and macro facilities are very helpful to an expert user.

3 Offer informative feedback.
For every operator action, there should be some system feedback. For frequent and minor actions, the response can be modest, while for infrequent and major actions; the response should be more substantial.

4. Design dialog to yield closure
Sequences of actions should be organized into groups with a beginning, middle, and end. The informative feedback at the completion of a group of actions gives the operators the satisfaction of accomplishment, a sense of relief, the signal to drop contingency plans and options from their minds, and an indication that the way is clear to prepare for the next group of actions.

5 Offer simple error handling.
As much as possible, design the system so the user cannot make a serious error. If an error is made, the system should be able to detect the error and offer simple, comprehensible mechanisms for handling the error.

6 Permit easy reversal of actions.
This feature relieves anxiety, since the user knows that errors can be undone; it thus encourages exploration of unfamiliar options. The units of reversibility may be a single action, a data entry, or a complete group of actions.

7 Support internal locus of control.
Experienced operators strongly desire the sense that they are in charge of the system and that the system responds to their actions. Design the system to make users the initiators of actions rather than the responders.

8 Reduce short-term memory load.
The limitation of human information processing in short-term memory requires that displays be kept simple, multiple page displays be consolidated, window-motion frequency be reduced, and sufficient training time be allotted for codes, mnemonics, and sequences of actions.

3.3.3 Heuristic guidelines

When Normans design principles are used in practice they are often referred to as heuristics. The term indicates that they should be adapted and applied to given design problems. They should be interpreted using past experience with focus on the

design context. Nielsen and Mack (1994) and his colleagues developed the ten usability principles below.

1. *Visibility of system status*
Always keep users informed about what is going on, through providing appropriate feedback within reasonable time.

2 *Match between system and real world*.
Speak the users' language, using word, phrases and concepts familiar to the user, rather than system-oriented terms.

3. *User control and freedom*
Provide ways of allowing users to easily escape from places they unexpectedly find themselves, by using clearly marked 'emergency exits'.

4. *Consistency and standards*
Avoid making the users wonder whether different words, situations, or actions mean the same thing.

5 *Help users recognize, diagnose and recover from errors.*
Use plain language to describe the nature of the problem and suggest a way of solving it.

6. *Error prevention*
Where possible prevent error occurring in the first place.

7. *Recognition rather than recall*
Make objects, actions, and options visible.

8. *Flexibility and efficiency of use*
Provide accelerators that are invisible to novice users, but allow more experienced users to carry out tasks more quickly.

9. *Aesthetic and minimalist design*
Avoid using information that is irrelevant or rarely needed.

10. *Help and documentation*
Provide information that can be easily searched and provides help in a set of concrete steps that can easily be followed.

3.3.4 User experience goals

In excess to the clearly designed usability goals there are less clearly defined goals for user experience. In many cases they might be hard to measure and vary greatly form user to user but they never the less give a good indication of how the user feels about the system. The ultimate goal for a system should be that it should be efficient and productive as well as pleasing and enjoyable for the user. Examples of user experience goals, according to Preece et al (2002), are that a system should be:

- Satisfying
- Enjoyable
- Fun
- Entertaining
- Helpful
- Motivating
- Aesthetically pleasing
- Supportive of creativity
- Rewarding
- Emotionally fulfilling

3.4 Usability Evaluation

It is important to evaluate the usability of a system to determine if it measures up to the usability goals and expectations that have been defined for the system. The usability evaluation results can be used to further improve the system or to compare it with other systems from a usability standpoint. There are several different methods of usability evaluation that all have different benefits and drawbacks. The following passages describe some of these techniques.

3.4.1 Field studies/background studies

In order for usability evaluation to be effective there is a need for a thorough background study. The background study is used to define and understand the user groups and to get a better understanding of the system that is to be evaluated and the context that it is used in.

Interview techniques

There are several different types of interviewing techniques (Ekholm, M., Fransson, A. 1992). They can be categorized into two pairs of categorizes; *Direct – indirect* and *highly structured – unstructured* ways of collecting information.

The direct way is to collect the information through the interviewer's own experiences during the interviews. The indirect way is to review material that is already collected.
The difference between highly structured and unstructured interviews is how hard the interviewer controls different situations that can affect the information being collected.

The type of technique that should be used depends highly on the situation. If the evaluating staff has little or no knowledge of the domain, a more indirect approach is probably the only way. This technique might feel a bit random at first, but the further into the interview, the more specific questions can be asked. Both a highly structured and an unstructured approach can be used to get an overview of the domain. Multiple-choice surveys are a good example of an indirect and highly structured interviewing technique.

Ethnographic observation and user analysis
According to Ottersten and Bajic (2004), the best way to conduct a user analysis is through studying how different users perform their work as well as how specific tasks are conducted. Based on that, study groups of users with common patterns of usage can be identified. The patterns of usage can be described as the sum of which tasks that are solved with the application, the experience and expectations of the user and the context in which the application is used.

Ottersten and Bajic argue that user groups rarely can be defined through grouping the users based on their position within the organization, the customer segment they belong too, or any other present definition of the users. The definition of user groups has to be based on the way that they actually use the product and this aspect can vary greatly between users, that for example have the same title within an organization.

An effective way to conduct a user analysis is through the combination of observations and interviews with the users. The interviews provide insight to the users opinions and experiences with the system although it often misses many aspects of how the system is actually used. An interview does not reveal how the system is used, only how the user thinks that he or she is utilizing the system. Because of this fact observations provide a good complement to interviews since they enable the observer to understand how the user actually utilizes the system.

A final aspect of the user analysis is prioritizing the user groups. The prioritization should be made based on how much benefit the user group in provides to the organization.

Surveys
Surveys can be used at any time in the life cycle, but are most often used in the early stages to better understand the potential user. Surveys are a very flexible tool. They can be shaped to investigate more extensive research and they can also be shaped to get a quick and easy interpreted view of standard questions.

The results from a survey have to be evaluated by a person with an insight in what kind of study the results are coming from. If the study includes more shallow and quick questions the result can easily be converted into graphs and statistics.

Survey questions have to be carefully picked out in order to not influence the survey participants. It is important to ask neutral questions with no personal opinions from the survey designers.

In SUS - A quick and dirty usability scale (Brooke, 1986), a system usability scale is discussed. This survey technique is used to gain a quick overview over the usability and attitude towards usability in a system. It consists of ten questions with a five-graded scale that asks direct questions regarding the user's feelings about the system. The result from a system usability scale (SUS) survey is, when finished and reviewed, a single number representing a composite of the overall attitude towards the system. Note that scores for individual items are not meaningful on their own.

The SUS scores have a range of zero to 100. To calculate the SUS score, first sum the score contributions from each item. Each item's score contribution will range between zero and four. For item one, three, five, seven and nine the score contribution is the scale position minus one. For items two, four, six, eight and ten, the contribution is five minus the scale position. Multiply the sum of the scores by 2.5 to obtain the overall value of SU. An example of the survey can be found in Appendix A.a.

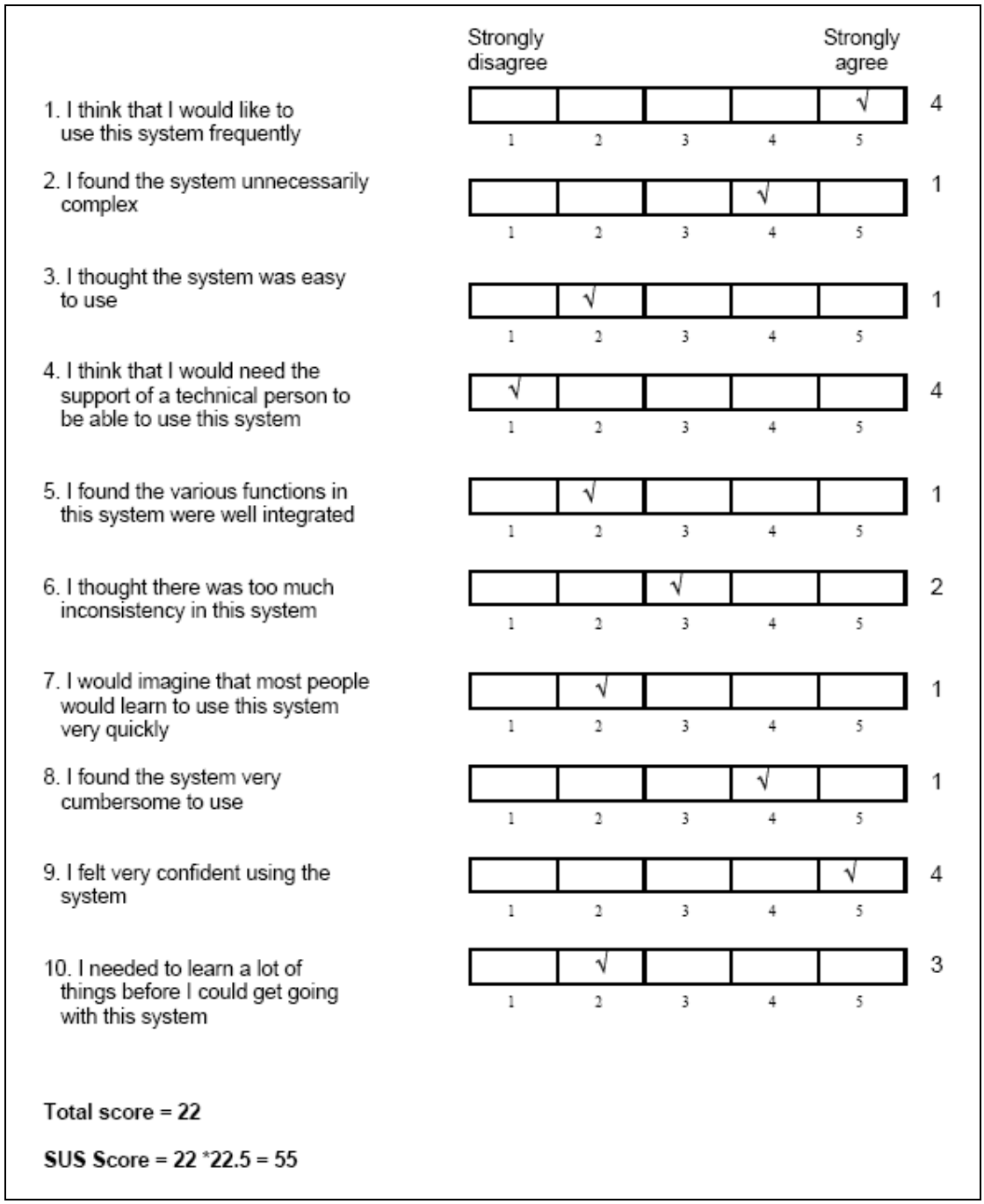

Figure 3.4 - System usability scale

3.4.2 Inspection methods

Usability inspection is a term that is used for several methods, some of the most important according to Nielsen and Mack (1994) are:

Heuristic evaluation - Usability specialists check whether each dialogue of a user interface obeys the rules of a small number of established usability principles.

Guideline review - An inspection is made of whether an interface meets the demands of a substantial amount of usability guidelines. There have been many attempts producing precise guideline collections but the final results often include approximately a thousand guidelines.

Cognitive walkthrough - A detailed procedure is used to simulate a user's problem solving process in every dialogue of the user interface to see if the next correct action is intuitive.

Pluralistic walkthrough - Users, developers and experts on human factors together step through a scenario, discussing usability issues associated with dialogue elements involved in the scenario steps.

Formal usability inspection - The formal usability inspection is a six-step process where several evaluators with explicit responsibilities first review the interface individually then merge their lists of usability problems.

Consistency inspection - System designers involved in other developing projects inspects an interface to check the consistency with their own design.

Standards inspection - An expert of specific interface standards inspects the systems' interface to check that it aligns to the standards of other systems on the market.

Feature inspection - Focus is set on the function in the software system, e.g. whether the designed function meets the needs of intended end users.

An additional method that is not described in Nielsen's and Mack's most important list is **Flow analysis**. This method can be used to support the cognitive walkthrough by illustrating a few important user scenarios in flow diagrams. These illustrations can be examined and evaluated to find usability flaws and problems with the user scenarios

3.4.3 Usability testing

The overall goal of usability testing is identify problem areas, within a computer-based interface, regarding user interaction and user satisfaction. There are several different ways of conducting a usability test and there are different pros and cons with every each and one of them.

Quick and dirty
Quick and dirty usability testing is an inexpensive way to test with real users in a short amount of time. It emphasizes on speed instead of formalness. Quick demonstrations followed by questionnaires are a great way of getting fast information regarding a product.

Predictive evaluation
Usability experts make the predictive evaluation. The evaluation itself does not require any users to be present at all. The usability experts, preferably more than three (Nielsen, Mack, 1994), to get good enough validity, do the evaluation based on theoretical principles, such as heuristic and cognitive walk-throughs. The experts

base the evaluation on their own experience and knowledge making the result better the more experienced the experts are. Often experts must analyze the results too, in order to make the most out of it.

Exploratory test
The exploratory test is conducted quite early in the development cycle, when the product is still in the preliminary stages of being defined and designed.

The idea of the exploratory testing is to examine or to explore the effectiveness of preliminary design concepts. These concepts are also known as the user's conceptual and mental model of the product.

Exploratory tests usually require close interaction between the participants and the test monitor. This is required in order to establish the efficacy of preliminary design concepts.

An exploratory test includes extensive prototyping and redesigning of the product. Prototyping can even be done with simple paper mock-ups or by designing it in the computer. It should be a fast way of realizing the new design ideas. The testing itself can be executed through cognitive walkthroughs.

Assessment test
Assessment test is probably the most commonly used usability test method. It is the most simplistic and easygoing method to both execute and design. It is mostly performed midway through a project after the fundamental design of the product has been established.

The purpose of the assessment test is to expand the findings from the exploratory test. The findings from the exploratory test were found to form a first conception of the system. Assessment test builds on them and creates a more in-depth understanding of the system.

The testing process is a mix of the loose exploratory test and the more strict and controlled validation test. In an assessment test the user always performs tasks, rather then just exploring the system. The test monitor takes on a more laid-back role and will not participate in the testing just as much. There will also be a chance to gather more quantitative measures with assessment test.

Validation test
The validation test is performed late in the development cycle and is intended to certify a product's usability. Therefore the test is executed much closer to the product's release than the two prior test methods.

The main objective of the validation test is to ensure that the final product fulfils the predetermined usability goals set up in the specification. The goals are typically stated in terms of performance criteria, such as speed and accuracy.

Another major objective of the validation test is to evaluate the product. For the first time it is possible to evaluate the product as a whole; how the different parts interact and how well software and hardware are integrated i.e.

Still another objective is to use the validation test as "disaster insurance". A company does not want to release a product containing major flaws and errors, having to send out fixes to every user; or even having to recall the product.

To conduct a validation test it has to be decided how adherence to the standard would be measured. Once this is done testing can begin.

Comparison test
The comparison test can be executed at any time during the development cycle. In the early stages it can be used to compare different prototypes to discard the least appropriate. In the middle of a project the comparison test can be used to measure the effectiveness of a single element. Towards the end of the life cycle comparison testing can be used to see how the product matches up to competitor's products.
The main purpose of the comparison test is to compare two or more designs. Comparison testing can, with advantages, be used in conjunction with any of the three prior methods.

The same tests are carried out on two or more versions of the product in order to collect the material needed to make the comparison. Once collected the test results are pretty straightforward and easy to compare.

3.4.4 Discount usability

According to research by Jakob Nielsen (1994), extensive usability engineering methods are rarely used in real life software development projects. He argues that even basic usability engineering principles such as early focus on the user, empirical measurement and iterative design, rarely get applied.

Nielsen proposes an alternative to the approach of using the best possible usability evaluation methods. He suggests that it is better to apply simplified and less time-consuming methods to ensure that usability work is performed. The suggested alternative is the "discount usability engineering" method that is supposed to be faster, cheaper to apply and often fit the timeframe of software projects better.

The "discount usability engineering" method is based on the use of the following three techniques:

- **Scenarios** – Performed by limiting the evaluated part of an interface by reducing the tested functionality. This is achieved through choosing key user scenarios that are to be evaluated. The scenarios can limit functionality in two ways: horizontally by limiting level of functionality and vertically through reducing the number of features.
- **Simplified thinking aloud** – A simplified form of user testing where the users are asked to think out loud while they perform tasks. This method is conducted without a usability lab and video equipment in order to save resources and time instead the test leader observes the user and takes notes. It is recommended that three to five users be tested to achieve the highest cost-benefit ratio of the usability testing.
- **Heuristic evaluation** – Many interface standards and collections of usability guidelines are far to extensive containing up to a thousand rules. These collections can therefore be very hard to follow and very time-consuming to use in a usability evaluation. An alternative is therefore to use a small set of usability principles as well as knowledge and experience while evaluating an interface.

3.5 Inspection vs. testing

In the book Usability Inspection Methods Nielsen and Mack discuss the differences between usability inspections and empirical testing. They claim that there is

economical value in replacing empirical evaluation with inspection methods, but at the same time there are a few trade-offs.

Usability inspections uncover different types of problems than empirical testing does. They are often not as complete as with empirical testing or they represent a different kind of problem.

Empirical testing is very popular because “seeing is believing”. Software developers often have difficulties believing there is a problem with their own software, until they see it on videotape. Empirical testing is a method that suits a project with the proper facilities and a more generous testing budget. Proper facilities can include a usability laboratory or the proper educated staff. The result from an empirical testing session is a reliable and valid result, if the session is performed correctly. The result can also be used to compare the system, being tested, with other similar systems. The set of results is often presented in tables and more exact numbers than what could be derived from an inspection. The biggest issue with empirical testing is the cost though. Empirical methods are often expensive and time-consuming, which gives inspection methods an advantage.

Inspection methods can, with favor, be used in projects with a limited testing budget and limited time. The amount of errors produced might not be as many as with empirical testing, but seen to the cost, inspections methods can well replace empirical testing to some extent. Given a certain time limit, inspections can be run parallel by different evaluators and can thereby be made more time efficient. The results from the parallel inspections can afterwards be collected and put together, making it a more reliable result. The more evaluators, the more reliable result. The biggest problem with inspections is that the validity and reliability is sometimes hard to prove.

3.6 Severity rating

Nielsen (1994) claims that in addition to locating usability problems in an interface, a heuristic evaluation can be used to rate the severity of the usability problems that were found. The severity ratings can be used to prioritize the problems according to how serious they are and how critical it is that they are fixed. The ratings also provide an estimate of the need for additional usability efforts. In the following passage Nielsen describes which factors that should be considered when rating the usability problems.

The severity of a usability problem is a combination of three factors (Nielsen, 2006):

- The *frequency* with which the problem occurs: Is it common or rare?
- The *impact* of the problem if it occurs: Will it be easy or difficult for the users to overcome?
- The *persistence* of the problem: Is it a one-time problem that users can overcome once they know about it or will users repeatedly be bothered by the problem?

The following zero to four rating scale can be used to rate the severity of usability problems:

- **0** = I don't agree that this is a usability problem at all
- **1** = Cosmetic problem only: need not be fixed unless extra time is available on project
- **2** = Minor usability problem: fixing this should be given low priority
- **3** = Major usability problem: important to fix, so should be given high priority

- **4** = Usability catastrophe: imperative to fix this before product can be released

In his research Nielsen suggests that the severity ratings should be made through gathering all usability problems that have been found during evaluation and specifying them in a document. The problem document should then be distributed to all evaluators that have been involved so that they can rate the problems according to severity. The evaluators do not need to have access to the system during while rating it, the ratings should be made fairly quickly and in most cases the rating activity should take about 30 minutes for the evaluator.

According to Nielsen, the reliability of the severity ratings depend on the experience and knowledge of the evaluators but the number of evaluators who rate the system is the factor that has the strongest correlation to reliability. Reliability of severity estimates as a function of the number of evaluators has been the focus of a study by Nielsen. The following graph is based on the Spearman Brown formula for estimating the reliability of combined judgments from several evaluators.

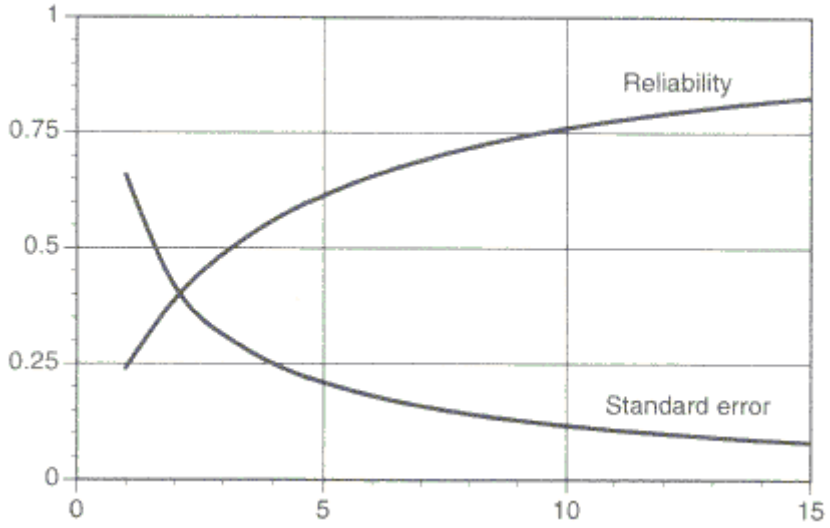

Figure 2.5 - Nbr of evaluators vs. quality

Based on this correlation Nielsen argues that the reliability of the severity ratings made by one evaluator is too low to base any major investments on. Nielsen further suggests that using ratings from three evaluators is enough for most practical purposes.

3.7 Cost-benefit analysis

According to Nielsen (1994), a cost-benefit analysis of usability evaluation and testing is made up of two elements: first the costs in terms of time spent evaluating are estimated, and after that the benefits in terms of increased usability are estimated usability (less the development costs for the redesign). The estimates involve some uncertainties and therefore they should be converted into currency using round numbers.

Nielsen further argues that the only way to get an exact measure of the benefits of a usability evaluation or test would be to fully implement two versions of the interface, one before the evaluation and one after. This would enable the evaluator to empirically test and measure the user performance. This method however is very

impractical and in many cases impossible to use since it would require implementation of the interface in many versions, and these versions would have to be used by a large number of users for a long time so that they reach expert performance levels. Nielsen goes on to discard yet another alternative which he describes as, “a detailed economic work-study model of the different steps involved in the users' workday in order to assess the frequency and duration of each sub-task”. According to Nielsen this approach is very detailed and time-consuming but not very accurate.

The model suggested by Nielsen is based on estimates of benefits and costs rather than measurement data, he argues that this approach, when properly applied, provides a fairly accurate result and can be used with great benefit. The strategy for cost-benefit analysis of a heuristic evaluation is described below, however this approach can be used for usability testing as well if the first steps that involve the locating of usability problems are substituted for testing.

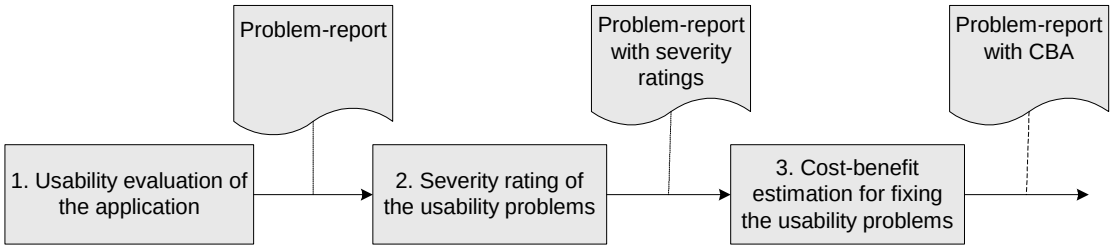

Figure 3.6 - CBA model described by Nielsen (1994). Authors’ interpretation.

Nielsen’s cost-benefit of heuristic evaluation consists of the following steps:

1. Evaluators (a number of usability experts, 11 in the study Nielsen has based his argument on) locate usability problems through heuristic evaluation.
2. The usability problems found by the evaluators are gathered and specified in one problem report.
3. The problem report is distributed to the evaluators who rate the severity of the usability problems.
4. The severity ratings are gathered, evaluated and specified into the problem report.
5. The problem report is distributed to the evaluators once again and they are asked to estimate the improvements in usability from fixing all the usability problems identified by the heuristic evaluation. Usability improvements are estimated with respect to two usability parameters:
    a. *Reduction of learning time*: How much less time would the users need to spend learning to use the system? Learning time considered as a usability parameter represents a one-time loss of productive time for each new user to learn the system, so any savings would be realized only once.
    b. *Speedup in expert performance:* Once the users have reached a steady state of expert performance, how much faster would they be able to perform their work when using a system with all the usability problems fixed than when using a system with all the problems still in place? Expert performance considered as a usability parameter represents a continued advantage for the use of the improved interface, so any savings would be realized throughout the lifetime of the system.

Nielsen claims that other usability parameters of interest include frequency of user errors and the users' subjective satisfaction, however these parameters were not estimated in his study and therefore are not included in this model.

6. The benefit estimations are gathered, evaluated and specified into the problem report. The outliers in the estimations are removed and average of the saving estimations are calculated
7. Calculations are made based on the estimations using the following steps
   a. The time the users actually spent working with the system is not the same as the total time working with the task. Estimate the fraction of time spent with the system according to Nielsen" Studies of the users indicate that they will spend about 1/3 of their time doing other tasks, 1/3 of their time performing the task without operating the user interface, and 1/3 of their time actually operating the interface". Multiply the speedup in expert performance with the estimated fraction.
   b. Estimate how many people work with the system and thereby are affected by the usability problems. Multiply the saved time of learning by the number of users to get the total saved time. Estimate how much time they work per year and multiply this with the number of users and the speedup in performance to get total savings in time.
8. Estimate cost per time unit and multiply.
9. Estimate the fraction of usability problems that it will be possible to correct (Nielsen claims that it is usually about ½ of all problems). Multiply the total savings in learning time and speedup of performance with this fraction.
10. Subtract the cost for evaluating the system and the cost for fixing the problems.
11. The result is the cost-benefit of the heuristic evaluation. Remember that the savings are not cash flow they are avoidance of penalty.

3.8 Plan for evaluation of system usability

The starting point for this thesis was the work that Anna Vagelin had done in her master thesis "Usability evaluation of a system in use" (Vagelin, 2005). Her work at AstraZeneca resulted in a plan that is composed of three phases based on how much time and resources that are available for the evaluation. The three phases are (authors interpretation of Vagelin, 2005):

Phase 1 – Initialization

This phase is the most basic phase and should take about two to three days to complete. The activities of this phase are:

- **Start up meeting** - A meeting with the system owner to start up the process. System documentation should be made available prior to the meeting.
- **Field Study** - It is recommended that at least the usability expert conducts a field study to get to know the context of the systems use, the users' way of working and the actual use of the system.
- **Simple Heuristic Evaluation** – It is recommended that the usability expert carry out a quick heuristic evaluation to gain better understanding of the user interface.
- **Interviews** - Interviews should be held with representatives of all user groups and with both new and old users to get the whole spectra of opinions on the satisfaction of the system. Also the owner and the service and support team should be interviewed about their contentment with the system. A second purpose is to become informed on how much support users give each other.
- **Judgment day** - When the investigator has committed a field study and interviews with users, support and system owner, its time for the investigation

group to have another meeting and make a subjective judgment on how to proceed.

Phase 2 – Discount investigation
If the investigation group decides that the investigation needs to go further, a second phase should be entered. The second iteration is still a quick, budget version of an evaluation but it should be planned to go on a little longer than the initialization, tentatively one to two weeks. The activities of this phase are:

- **Start up meeting** - In the beginning of the second phase, a start up meeting should be held by the usability expert to clarify what the iteration involves.
- **Contextual Analysis and Process comparison preparation** - A contextual inquiry should be performed to find unarticulated aspects of user tasks.
- **Heuristic evaluation** – A more thorough evaluation, with two extra usability experts.
- **Analysis of user support calls** - The usability expert works together with the member of the service and support team to analyze the different kinds of user service requests.
- **Error rate measuring** - Suspected data errors should be measured, e.g. by logging actual use.
- **Judgment day** - The second judgment day involves discussion about the results so far, and if further investigation is needed.

Phase 3 - A thorough investigation
The third phase needs further planning by the usability expert. The elements below are merely a suggestion on what can be included. How they should be combined depends on the results of earlier phases. The length of the third phase may differ greatly, but a minimum of three weeks should be reserved. The activities of this phase are:

- **Start up meeting** - Another start up meeting where the usability expert clarifies what the iteration involves
- **Process comparison** – A discussion between the expert and system owner. If the system does not fit into the process, can the process be changed? If the system does fit into the process, is it clear to the user?
- **Usability testing in lab** - To understand how users actually interact with system interface, Usability Testing should be performed.
- **Analysis of first line support** - Questionnaires can be used to gather statistics on how much time users spend helping each other to be able to conduct their tasks.
- **Tracing effects and estimate costs** - Trace the effects of the usability problems found in the heuristic evaluation. The usability expert should pick the cases where the greatest impact is believed to exist and limit cost estimates to them.
- **Judgment day** - The usability expert should have produced a report on the results of the evaluation methods to work as basis for another decision on how to proceed. It is now time for the final judgment on the system; how should it be classified?

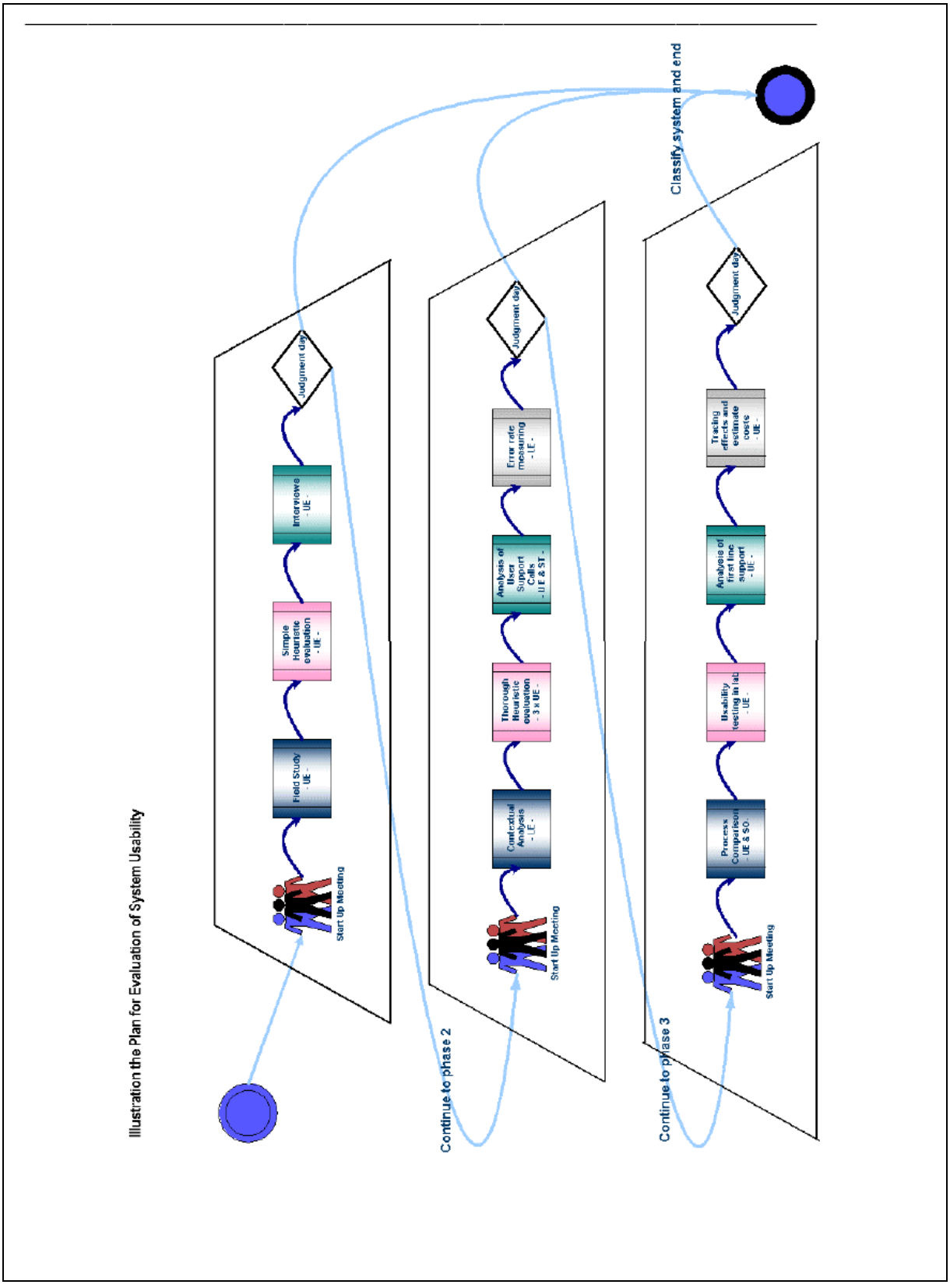

Figure 3.7 - Illustration of "Plan for evaluation of system usability"

4 First framework proposal

To satisfy the objectives of this thesis, a preliminary framework for usability evaluation was proposed. The preliminary version of the framework was developed with the purpose of testing the theories in a real life setting. This chapter describes the framework activities that were developed prior to the case studies, as well as motivates why they have been chosen for testing. The choice of activities will be motivated by, and connected to, the theory chapter. However, in this chapter the activities will not be described in detail. How the activities were applied is described in the case study chapters. The final framework will be discussed and described in detail in the chapters following the case studies.

The first version of the framework was divided into four phases:

- Background studies
- Usability Evaluation
- Usability severity rating
- Cost-benefit analysis

4.1 Background studies

The main goals of the background studies phase are that the evaluator acquires an understanding of, the system that is under evaluation and the roles of the staff surrounding it. A secondary goal is for the evaluator to form an opinion of the objectives that the system is supposed to fulfill, as well as the most critical user-tasks that the system is supposed to support. To achieve those objectives the following activities were proposed:

4.1.1 Start up meeting

The startup meeting is the first activity of the framework. At the startup meeting the usability staff meets with the stakeholders of the system that is to be evaluated. During the meeting, the goals of the evaluation are defined; the budget for the evaluation (time to spend) is also set. Contact persons who have responsibility for the system or certain system parts are assigned to the project to assist the evaluator.

4.1.2 Documentation studies

In order for the evaluator to understand the system, he/she needs to get an overview of the goals that the system is supposed to fulfill and the functions used in the system to satisfy these goals. A good way to do this is through studying system documentation and tutorials or courses on how the system should be used. Having an overall understanding of the system is a key factor if the evaluator is going to get the most out of the following evaluation activities.

4.1.3 Interviews

The objective of the interviews is to gather information about the system and the way it is used. Interviews are conducted with key stakeholders as well as different types of system users. In order for the evaluator to get the most out of the interviews a direct type of interview is recommended with a combination of a structured and unstructured approach. The combination of predefined questions and open discussion should enable the evaluator to gather as much information on the system as possible.

4.1.4 Field study/observation

The objective of the field study/observation is to gather information on who the users are, how the system is used and provide an insight to what opinions and experiences the users have with the system. According to Ottersten and Bajic (2004), the best way to conduct a user analysis is through studying how different users perform their work as well as how specific tasks are conducted. An effective way to conduct a user analysis is through the combination of observations and interviews with the users. In many cases the two can be combined by observing the user in action, and following up the session with an interview.

4.2 Evaluation

In the usability evaluation phase the system is studied based on the information that was gathered in the background studies phase. The objective of the evaluation phase is to find and specify usability problems that are present in the system in a usability problem report and gather attitudes towards the system. The evaluation can be divided into two separate activities, attitude surveys and usability evaluation.

4.2.1 Attitude survey

An important aspect of the usability of a system is the users attitude and opinions towards the system. In order for the evaluator to gather information about the user attitudes and opinions a survey was proposed. It consists of ten questions with a five-graded scale that asks direct questions regarding the user's feelings about the system. The result from a SUS survey is, when finished, reviewed and yields a single number representing a composite of the overall attitude towards the system

4.2.2 Usability evaluation

In addition to gathering the opinions and attitudes that the users have towards the system, an expert evaluation was proposed. The expert evaluation serves to find problems that the users have not thought about, as well as problems that they already pointed out but did not know the underlying reason for.

In the preliminary framework, a few methods were recommended for testing during the case studies. The first was a broad heuristic evaluation of the system based on Norman's *Design principles,* Nielsen's *eight golden rules of interface design* and Shneiderman's *Eight golden rules of interface design* which was recommended to find the parts of the interface that did not follow the conventions for good usability. The second was a deeper evaluation of the most critical user-tasks (that were defined through the background studies) based on the same principles. The critical user tasks should also be evaluated through a Flow analysis with respect to if they achieved the goals that they were supposed to, in an effective way.

A third part of the usability evaluation that was considered was the analysis of user support calls. The purpose of this activity is to analyze and find the problems that they users most frequently need help with in order to find the underlying usability problems. Finally to end the evaluation the interface should be checked towards specific usability guidelines that have been defined for interfaces at AZ.

4.2.3 Usability problem report

The results from the evaluation activities should be specified in a usability problem report. The problem report should contain a specification of the problem and were in the system it occurs. The problem report should summarize all the found problems, from the background studies through the usability evaluation.

4.3 Usability severity rating

The purpose of this phase is to prioritize the usability problems and to collect information to support the cost-benefit analysis for fixing the problems.
After the usability problems have been gathered the evaluator makes first prioritization of the problems by selecting a number of errors that he/she feels are the most critical. The number is based on the extent and budget of the evaluation. The selected problems will then be specified in a separate document, the usability problem report 2, for the purpose of a severity rating. The suggested severity rating method is based on the three factors suggested by Nielsen (2006):

- The *frequency* with which the problem occurs: Is it common or rare?
- The *persistence* of the problem: Is it a one-time problem that users can overcome once they know about it or will users repeatedly be bothered by the problem?
- The *impact* of the problem if it occurs: Will it be easy or difficult for the users to overcome?

In his research Nielsen suggests that the severity ratings should be made through gathering all usability problems that have been found during evaluation and specifying them in a document. The problem document should then be distributed to all evaluators that have been involved so that they can rate the problems according to severity.

Due to the limited timeframe for a usability evaluation using this framework a few changes were made to Nielsen's suggestion. Instead of distributing the document to several evaluators, the document should be distributed to a number of stakeholders that have insight into the system. The usability evaluator as well as the stakeholders then rates the problems separately. The suggested method for gathering the ratings is a web survey (Appendix A.a).

The answers to the web survey should then be summarized and entered into the second problem report.

4.4 Cost-benefit analysis

The cost-benefit analysis (CBA) phase has the purpose of calculating a potential monetary value for fixing the usability problems. The usability problems that have been found to have the highest severity rating should be chosen for analysis. The suggested method for CBA is for the evaluator to have a meeting to discuss the usability problems with a stakeholder who has extensive knowledge of the system and the administration around it, such as a system owner or a super-user. At this meeting the usability improvements are estimated with respect to two usability parameters:

1. Reduction of learning time
2. Speedup in expert performance

The input to the meeting is

- The usability problem report with severity ratings
- Statistics on how many users there are that are affected by each problem
- The cost per hour on average for these users

Based on that input the savings are estimated and the value is calculated.
The final activity in the framework is to specify the results into the final evaluation report. This report should also contain a summary and analysis of the attitude survey results.

5 Case study – StudyCapture

The StudyCapture case study was performed with the primary purpose of testing the preliminary framework for usability evaluation, which was presented earlier in this thesis. A secondary objective was to perform a usability evaluation of the StudyCapture system with the purpose of finding usability flaws.

5.1 System description - StudyCapture

StudyCapture is a web based data capture tool (WBDC) designed by AstraZeneca to capture clinical study data. It is used to support the clinical studies process by capturing data and making it available to the relevant functions and people within the organization. The system was introduced as a web-based system to substitute the old paper process. The system is based on the old procedure of communication at AZ in a clinical study. At the time of this thesis StudyCapture was still being used at AZ but a new system was being introduced to replace this one. The StudyCapture system is used by investigator staff that conduct the clinical studies (M.d. and nurses) as well as monitors (IT-staff) who control the entered data. There are other user groups as well but investigators and monitors make up a majority of the users.

5.2 Background studies

The main goals of the background studies phase were that the evaluator was to acquire an understanding of StudyCapture and the roles of the staff surrounding it. A secondary goal was for the evaluator to form an opinion of the objectives that the system was supposed to fulfill, as well as the most critical user-tasks that the system was supposed to support. To achieve those objectives the activities in the following paragraphs were performed.

5.2.1 Start up meeting

The evaluation process was initiated through discussions between the usability expert (referred to as the evaluator) and the staff members at AZ who had suggested the StudyCapture case study as a part of the evaluation of the framework. The purpose of the evaluation was to evaluate the preliminary framework and to perform a real evaluation of the StudyCapture system. It was agreed that the case study was to be spread out over a couple of months as each part of the framework was developed in parallel with the case study. It was also agreed that the evaluator was to partake in an e-learning course for the system and receive a short presentation of the organization around the StudyCapture system. Further more the discussions resulted in a list of other staff members that could be used as contacts throughout the evaluation.

5.2.2 StudyCapture e-learning

In order for the evaluator to gain a general understanding of the StudyCapture system he went through an e-learning course that is used to introduce the monitors and investigators to the system functionality. After the course the evaluator was able to understand the basic functionality of StudyCapture and the tasks that the investigators and monitors faced in their work.

5.2.3 Interview with super-user

An interview was conducted with a super-user of the StudyCapture tool. The purpose of the interview was for the evaluator to gain more insight into how StudyCapture was used and how the organization around the system was put together. A second purpose was to get information about the views and opinions that the super-user and staff that had reported to him, had about what was positive and negative with the StudyCapture system. The super-user possessed extensive knowledge about the system and had been involved in setting up and administrating studies. The super-user had also been involved in educating others in the use of the system.

The interview was conducted as a direct and semi-structured interview, were a few questions were prepared but an open discussion was encouraged.
The super-user was asked questions on the following topics:

- His background and role today
- The organization around StudyCapture
- Problem reporting and support in StudyCapture
- His views on the pros and cons of the system
- Which parts of the system those were critical to focus on during the evaluation.

After the interview the evaluator had gained a better understanding of StudyCapture and the organization surrounding it. The full results of the interview will not be presented here, however the most important results will be summarized below. The interview resulted in a few key facts about the StudyCapture organization, a list of usability problems, and a list of user tasks that the super-user felt were the most critical for the evaluation to focus on. The usability problems are listed in the first usability problem report (5.7 - *Usability problem report 1 – First summary*). The most important organizational facts for the purpose of the evaluation were:

- The support functions were the evaluators can turn to in order to find the most common problems were:
    A. For questions regarding technical problems the users turn to the technical support
    B. For questions regarding functionality the users ask other according to the following chain: investigator -> monitor -> central study team -> super-user/ISM (implementation support manager) -> technical support
- The central study teams should be virtually independent and therefore usability problems should rarely get reported to the super-users.
- The study-setup process can be difficult to administer since part of that process is still done on paper and is not integrated in the system.

The user tasks that the super-user felt were the most important for further evaluation were:

- The handling of serious adverse events, since there are strict regulations from government agencies regarding those reports.
- The process where monitors ask questions to the investigators through the system and the investigators have to answer them.
- The parts of the data input process were several items have to be entered and submitted (saving of data) were there is a risk of loosing the data if it is not submitted.

5.2.4 Observation and interview with monitors

The observation and interview with monitors was conducted with the intention of learning more about the actual users of the system. The objective was to gather

information about their background, working situation, opinions and use of the system. The evaluator accompanied two monitors to a monitoring visit at a care-center. Once there he observed their work and asked questions along the way. The interview was conducted throughout the visit and a number of set questions were used as a basis for an open discussion. One of the two monitors only had a few months experience with the system, while the other had been working with StudyCapture for several years.

The monitors were asked questions on the following topics:

- Their background and roles today
- The organization around StudyCapture
- Problem reporting and support in StudyCapture
- Their views on the pros and cons of the system
- Which parts of the system that were critical to focus on during the evaluation
- Their everyday work with the system

The observation and interview with the monitors enabled the evaluator to gain an understanding of how StudyCapture was used by two primary system-users. It became apparent which system functions that were most critical for these users and in what context they were working with the system. The opinions and views that the monitors had towards the system were also recorded. Some of the most important information that the evaluator required during the activity is presented below.

- The monitors work with StudyCapture at the clinical study sites as well as from their office.
- At the site it can often be in a cramped space and with limited time.
- The source data verification has to be made at the site since the information cannot be removed from there.
- The monitors seldom ask staff outside the study team for help. Questions stay within the group.
- Some details in StudyCapture can be adapted at the setup of each study but are much harder to change after the study-setup. Other details in the system cannot be changed.
- The monitors use a manual to aid them in their work process.

The user tasks that the monitors thought were the most critical to examine further are the following:

- The list of generated exceptions from the overnight data checks. It can be very hard for the monitors to translate the information into questions to the monitors.
- The data entry process since that is what the investigators spend most of their time with.
- The process of asking and answering questions in StudyCapture.

Other problems with the system that the monitors pointed out are translated into usability problems and reflected in the first usability problem report (5.7 - *Usability problem report 1 – First summary*).

5.3 Evaluation

In the usability evaluation phase the system was studied based on the information that was gathered in the background studies phase. The objective of the evaluation phase was to find and specify usability problems that were present in the system, in the usability problem report and to gather attitudes towards the system. The evaluation can be divided into two separate activities:

- Attitude surveys
- Usability evaluation

5.3.1 Attitude Survey

An important aspect of a usability evaluation is the attitude that the users have towards the system. The attitude aspect is not covered in the usability problem report, and therefore the attitudes were evaluated separately through an attitude survey. The survey that was used to gather the user attitudes towards StudyCapture was *SUS - A quick and dirty usability scale* (Brooke, 1986). A few extra questions were added before the SUS survey in order to collect some extra information for the evaluation The SUS survey was used to gain a quick overview over the usability and attitude towards usability in a StudyCapture. The SUS survey with the additional questions can be found in Appendix A.a. The scoring of the survey yields a result between zero and 100 where 50 is a neutral response. The survey was created and distributed as a web survey using a tool at AZ. The survey was distributed to users of the StudyCapture system.

Survey Results

28 staff members at AZ answered the survey. All of the staff members were also current users of StudyCapture. The StudyCapture roles of the 28 staff members that filled out the survey were distributed as follows:

- 14 staff members worked only as monitors.
- 10 staff members were not monitors. Examples of their roles were study coordinator, study team member, StudyCapture administrator and StudyCapture trainer.
- 4 staff members had both the monitoring role and another role such as study team leader or StudyCapture trainer.

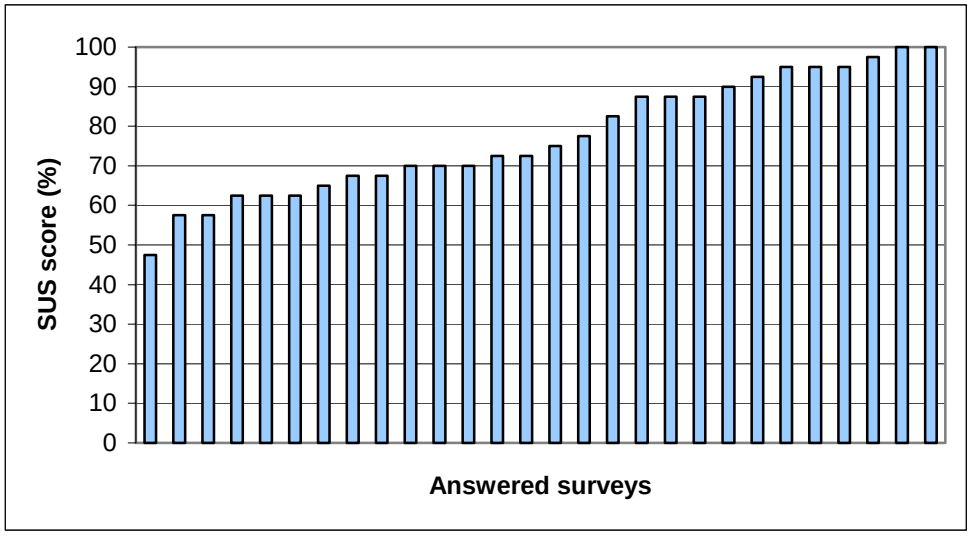

Figure 5.1 - SUS scores of all surveys

Table 5.1 – Average time using the system per week

Role:	Time:
All roles	9 hours and 6 minutes
Monitors	9 hours and 54 minutes
Non monitors	9 hours

The staff members spend on average 9 hours and 6 minutes with the system per week according to the survey. The average for staff with a role that includes monitoring the average is 9 hours and 54 minutes per week, and for staff that does not have a monitoring role the average is 9 hours exactly. Furthermore the staff members have been working with StudyCapture for an average of 23.5 months.

Table 5.2 - Total scores

Value:	Score:
Average	77.4
Median	73.5
Low	47.5
High	100

The average calculated score for the StudyCapture attitude survey is 77.4 and the median is 73.5. This means that the system has received a positive rating from the users that participate in the survey. Another notable fact is that only one person gave the system a rating of less than 50 (47.5), which shows that almost all of the users had a positive attitude towards the system.

Table 5.3 - Experience vs. scoring

Experience:	Score:
< 6months experience	73.75
>= 6months experience	78.5

If the scores for users with less than six months experience, are compared to the scores from those that have more than six months of experience the results showed that there is a scoring difference of 5 steps in favor of those with more experience. This indicates that users with more experience have a more favorable attitude towards StudyCapture than those with less experience. This could be interpreted as that the long time users have taken a better liking to the system and therefore give it a better rating then those that have not had the opportunity to get as familiar with it. It might also be an indication that it takes a while to understand and get a feel for the system.

Table 5.4 – Hours per week vs. scoring

Hours per week:	Score:
< 5 hours per week	73
>= 5 hours per week	79.5

If the data is further analyzed a comparison between users who utilize the StudyCapture for five or more hours/week can be compared to those that use the system less than five hours/week. The users that spend more time with the system give it a 6.5 steps higher rating. This might once again indicates that more experience with the system results in a higher rating. Another conclusion that can be

draw is that perhaps users forget how to use some functions between each usage and therefore give it a lower score if the use is not as frequent.

5.3.2 Usability evaluation

The usability evaluation was performed in four steps:

- A general heuristic evaluation of the StudyCapture as a whole.
- A deeper evaluation of a few specific user tasks that were defined based on the background studies
- The analysis of the problems based on support calls.
- The writing of a usability problem report that summarizes all the usability problems that were found in the StudyCapture system.

To start of with a heuristic evaluation of the system based on Norman's *Design principles,* Nielsen's *Heuristic guidelines* and Shneiderman's *Eight golden rules of interface design* which was performed to find the parts of the interface that did not follow the conventions for good usability. The evaluation was made in a training version of StudyCapture that AZ supplied to the evaluator. The evaluation was performed in two parts. First the system was evaluated as a whole and after that it was evaluated based on the critical user tasks that were derived form the background studies. The critical system functions that were studied extra carefully were the following:

- The data entry process since that is what the investigators spend most of their time with.
- The process of asking and answering questions in StudyCapture.
- The parts of the data input process were several items have to be entered and submitted (saving of data) were there is a risk of loosing the data if it is not submitted.
- The handling of serious adverse events
- The process where monitors ask questions to the investigators through the system and the investigators have to answer them.
- The list of generated exceptions from the overnight data checks. It can be very hard for the monitors to translate the information into questions to the monitors.

In addition to heuristically evaluating the critical functions, the functions where evaluated through a flow analysis to check whether the course of actions for completing that task were logical and fulfilled their purpose effectively.

After the heuristic evaluation was performed the information that was gathered in the background studies phase was analyzed and converted into usability problems. The information that was used for this purpose was a mixture of the opinions and views of individual users and a summary of problems that were reported to various support functions such as the technical support and informal support between users.

The final activity of the evaluation was to summarize and specify all the usability problems that were found in the system. The problems were described in the usability problem report 1 (5.7 – Usability problem report 1 – first summary); a report that was to be used for severity rating of the problems. The report contains a description of each problem and a description of the possible consequences from it.

5.4 Usability severity rating

The purpose of the usability severity rating face was to prioritize and rate the usability problems that were found and specified in the prior phases of the framework. The results from this face will then be used to conduct a cost-benefit analysis for fixing the problems. The evaluator, based on an estimate of their severity, selected eleven problems from the usability problem report 1. The selected problems were more thoroughly described in a second usability report that was to be used for the usability rating. The problems were rated based on the frequency and the impact of each problem.

The rating was conducted through distributing the document with the eleven usability problems, along with a web-survey with a set of questions corresponding to each problem. In the web-survey, the ratings for each usability problem were recorded. The severity rating is based on the following questions and a set scale for each question:

- How frequent do you think that the functionality that is affected by the usability problem is used by the user group it applies to?
- Out of the times that the functionality is used, how often do you think that this usability problem affects the user in his/her work with the system?
- How serious do you think that the possible consequences of this usability problem are?

The full web-survey questions can be read in Appendix A.a.

Severity rating results

The severity rating survey was distributed to three staff members; only one of them, a monitor, answered the survey. In addition to the one answer the evaluator also filled in the survey. The results from the survey were calculated and summarized for each problem. An analysis of the severity rating enabled the evaluator to prioritize the usability problems according to their rating.

Table 5.5 - Prioritization of usability problems

Problem:	Severity rating:
Problem 11	21
Problem 3	18
Problem 9	17.5
Problem 10	14
Problem 5	9.375
Problem 4	5
Problem 6	5
Problem 7	5
Problem 8	5
Problem 1	3
Problem 2	1.875

The severity rating and the average results from each question in the web-survey are presented along with each usability problem in the usability problem report 2. Based on the prioritization of the usability problems the five problems with the highest severity rating were selected to be further analyzed through a cost-benefit analysis.

5.5 Cost-benefit analysis

A cost-benefit analysis (CBA) for the most severe usability problems was performed with the purpose of calculating the benefit market for fixing the problems. The input to the CBA was a problem report with the five usability problems that had received the highest severity rating and additional information about the system and it's users.

The CBA was conducted through a meeting were each problem was discussed, and the costs that each problem inflicts on the company were estimated. The evaluator led the meeting; the other participants were two staff members with administrative roles and an extensive knowledge on the StudyCapture system. One of the staff members had prior experience and expertise on CBA models and was therefore able to give input on how to perform the analysis.

For each problem the frequency at which it occurs and how much time that has to be spent on its consequences were estimated. The cost benefit analysis that was performed in this case study was not complete since the costs for fixing the usability problems and the savings that the changes resulted in were not investigated. The analysis in this case study calculated the benefit-market for each problem. The benefit-market for a problem is defined as the maximum savings that are possible if the problem was eliminated completely through a perfect solution. The resulting benefit-market gives an indication of whether each problem should be investigated further and if there potentially are savings to be made. However the benefit-market does not indicate how large the actual savings will be, the actual savings can only be calculated after the developers of the application have presented an alternative solution.

During the meeting the possible consequences of each problem were defined in a diagram and then the frequency of each consequence, as well as the time it consumes, was estimated. The estimates were calculated in "full time equivalents" (FTE). One FTE is the equivalent of on staff member working full time. The following formula was used for calculating the FTEs:

*(problem min/day/user * nbr of users) / (total work minutes/day) = number of FTEs*

The complete results of the analysis of each problem are presented in usability problem report 3 along with an explanation of how the results were reached. A short summary of the results is presented below.

The following table displays predicted benefit market for each problem. For each problem the expected cost and the minimum cost was estimated. The calculations are made with the input that there are 771 users (monitors) who are affected by the problems. There might be a higher number of users that are actually affected but the only users that the benefit is calculated for is the monitors.

Table 5.6 -The benefit market for each problem

	Average costs in FTEs	Minimum costs in FTEs	Notes
Problem 11	14.4	4.8	
Problem 3	1.28	0.7	
Problem 9	50.2	33.5	Not a usability problem
Problem 10	?	?	Not a usability problem
Problem 5	0.30	0.15	

The costs in FTEs associated with each problem are an indication to if the problem should be further explored. The summary table above indicates that for problem 11, 3 and 5, a further investigation would be motivated. For problems 9 and 10 a further investigation might also have been motivated, however during the CBA meeting an agreement was reached that these problems were not usability problems. Problems 9 and 10 were problem that required a change in functionality and therefore should not be evaluated as usability problems.

To complete the CBA of the usability evaluation the developers of StudyAdmin would have to come up with a suggestion on how to improve the system and fix the specified problems. The new solutions would then have to be compared to the current to investigate how much the costs could be reduced. There would also have to be a summary of the costs associated with this evaluation such as the salary for the evaluator and the other staff members who were involved in the evaluation. The final figures could then be summarized in a table according to table 5.7.

Table 5.7 - Summarized CBA calculation

Total cost reduction (+)	
Costs for fixing the problems (-)	
Evaluation costs (-)	
Total benefit (=)	

5.6 Case discussion

As the StudyCapture was performed with the main objective of testing the first proposal for the usability evaluation framework the results may not be structured as, or match the results that would come out of the use of the final framework proposal. Since each step of the framework was developed in parallel to the case study the evaluation took a lot longer than the 40 man-hours that were defined as an objective for the framework. Furthermore, since the main objective was to test the framework, the validity and reliability of the results of the case study will not be analyzed in depth for this case study. However a short discussion of the results of each framework phase will be discussed below.

Background studies

The first phase of the case study served the purpose to provide a background for the evaluation. Since the framework was not defined at the start of the project no formal start-up meeting was held were the scope and goals of the evaluation were defined. Instead the scope of this entire thesis was defined, however the key staff members that would aid in the case study were provided. The semi-structured interviews with the monitors and the super user supplied the evaluator with the information that was needed to continue with the evaluation. Along with the education in StudyCapture and the observation of how the system was used a solid foundation for the evaluation was accomplished. All the activities mentioned above could be improved, but the information that came out of the activities was considered trustworthy and fulfilled the purpose.

Evaluation

28 staff members filled out the attitude survey. Through the answers the evaluator was able to draw some conclusion that the overall attitude towards the system was positive. One flaw of the attitude survey was that since it had not been used at AZ before no conclusions on how the results could be compared to those from other

evaluations could be drawn. The comparisons experience-scoring and frequency of work scoring also need to be confirmed further to be trustworthy. Another conclusion is that the extra questions that were added to the survey enabled a more extensive and detailed interpretation of the results. The analysis of the results could have been improved if the selection of staff members was more thought through. For example it would have helped if the staff members were more evenly divided according to roles and experience.

The general heuristic evaluation produced extensive results. The more detailed evaluation of the critical user tasks also enabled the evaluator the find process specific problems. On the background of the CBA meeting the critical tasks should have been selected differently as it became apparent that two of the tasks were selected based on problems that were not usability related. The heuristic evaluation that was performed and the problem report that was written were a lot more time-consuming than they could have been if the timeframe of 40 man-hours for the entire evaluation was held. The results of the evaluation could have been improved if an evaluator conducted the evaluation with more experience in heuristic evaluations.

Usability severity rating
The severity rating survey was sent distributed to three staff members and the evaluator. One staff member and the evaluator only answered the survey. The small number of answers gives the severity ratings a low reliability. Distributing them to more users who have different roles when working with the system would improve the reliability.

Cost-benefit analysis
The approach to the cost-benefit analysis change during the CBA meeting, since the meeting was used as an opportunity to give critique on the process. Since the focus was on coming up with the best approach for a CBA the results were not analyzed as thorough as could have. During the CBA meeting two the attendees concluded that two of the problems were not usability problems. This realization showed that the problem selection process could be improved but it also showed that the meeting between system experts and the usability evaluator was a good activity for a final prioritization of the problems. At the CBA meeting the reduction of learning time for each problem was not estimated, partly because the StudyCapture was about to be substituted for another system and partly because there was missing information on what could actually be considered as learning time.

5.7 Usability problem report 1 – First summary

This report contains the problems that were found in the StudyCapture application. The report was produced as a part of a case study with the aim of developing a framework for usability evaluation and cost-benefit estimation of fixing usability problems. The problems in this report were found through going through the first steps of the preliminary framework. The activities that were performed to find the problems were:

- Literature and documentation studies
- An observation of the system in use
- Interviews with key users and stakeholders associated with the system.
- Usability evaluation
- The first part of this report describes the general usability problems that were found throughout the system. The second part describes the problems that were found when the focus of the evaluation was on a few key tasks that

were defined through the background studies and interviews with the system stakeholders.

5.7.1 General problems

This section contains usability problems that were found during the usability evaluation. The problems that are specified in this section are problems that either could not be connected to a specific key task or were problems that could be generalized and regard the entire system.

Help Functions

1. If the user needs help with a specific part of the process he or she needs to find it in the help section. This makes it harder and more time-consuming to find than if it is connected to the screen that is viewed at the time help is requested.
2. On several data entry screens there is a button, which is labeled "Instructions". When this button is pressed a message states "No instructions found for this module." This means that there is no reason for the button to be present.

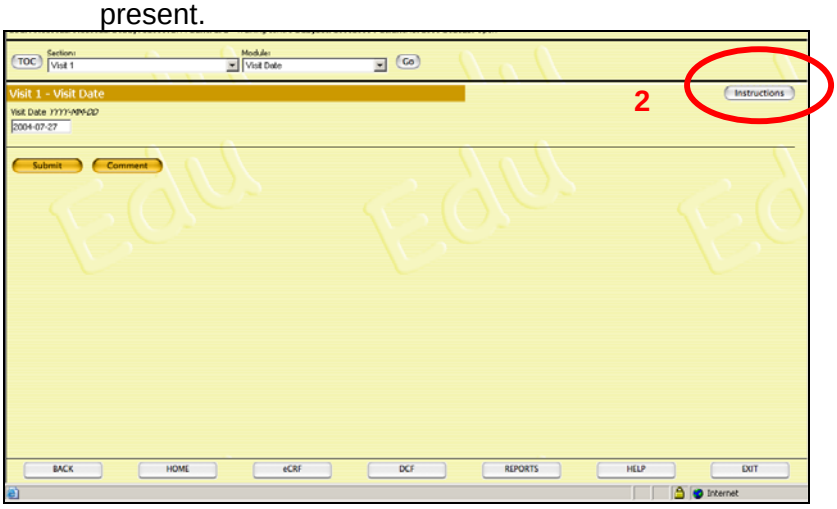

Figure 5.2 - Visit 1, visit date

Updates

On some occasions there are problems with the sharing of resources when several users are handling the same data. If an update is made by an investigator while a monitor is checking the same data, the monitor might have to exit the module and go back in to be able to see the update.

System orientation and Navigation

1. There should not be a label called "home" or a label called "TOC" they should refer to the page by name. Structural flaw enable you to reach the same place through different links. This makes it extremely hard to learn how to use the system.

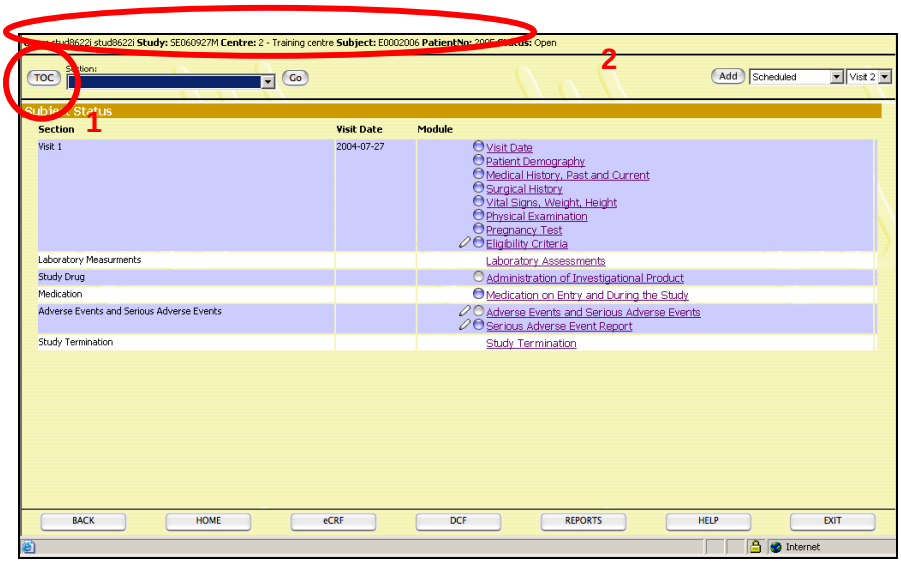

Figure 5.3 - eCRF

2. The system is using breadcrumbs to help system orientation. However they are not used consistently. This functionality can be improved by using breadcrumbs for all levels and to use them consistently. If they were made into links it would help the user get a sense of location and orientation throughout the system. In combination with different ways to reach different system states confuses the user.
3. There is no distinction of what level in the system you are at. It makes it hard to distinguish between entry and overview pages. Use colors, or other distinctions to distinguish different levels and functions in the system. This should be done carefully and with minor variations of color, for example.

Descriptions

There is not always a natural mapping between descriptions and the functions they describe throughout the system. The descriptions should be connected to their function through some type of grouping.

Messages

1. Error messages need to be connected to the fields or functions they relate to and only specify what is important for the user to know.
2. The use of feedback text in red is questionable when an action was successful. Convention is that red indicates failure. Also the feedback text is not grouped with the action taken.

Terminology

It is not recommended that abbreviations be used in an interface unless they are absolute convention. In cases when the meaning is not clear there should be a description. There also needs to be a list of accepted abbreviations for input that users can refer to and find easily.

Visibility and Color

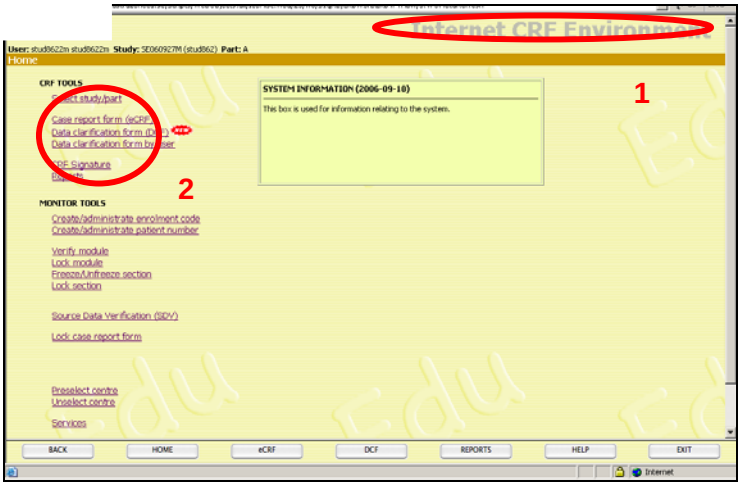

Figure 5.4 - Start page monitor

1. On several occasions the color contrast between the text and the background could be better. Examples of this are:
    a. The color of the text “Internet Crf Environment” should have more contrast to the background to improve readability. Blue-black contrast on tables
    b. The color contrast black-blue in several tables.

   Improving the contrast would improve the readability of the interface.
2. The color purple is in many cases a convention for a clicked link and indicates that it has been visited. This might fool the user into thinking that the section has been visited.
3. Italic text in the interface is not emphasized enough, in areas where several choices are possible an italic text specifies if several alternatives are to be chosen or if the user should pick only one. In that scenario the possibilities are indicated by an “or” or an “and” between the choices. If the text is not emphasized better the user might mark to few or to many alternatives, which will result in extra time to correct errors.

Text Boxes

1. The text boxes that are not applicable should have constraints on them that indicate that nothing should be entered into them. For example if you have chosen that the subject has not undergone any relevant surgery, you should not be able to enter data into the fields were surgery is specified. Boxes that do not need data should be locked.
2. The field labels should be more descriptive of their function. Many times in the interface, the labels are ambiguous. It could help if the study setup people had more insight into the knowledge of the users. The terminology also needs to be more adapted to the users.

Visibility – Resolution

Try to avoid scrolling down to find things as much as possible. Scrolling down to find links and text makes it harder to find it. Use the space that is available to avoid hard-to-find links.

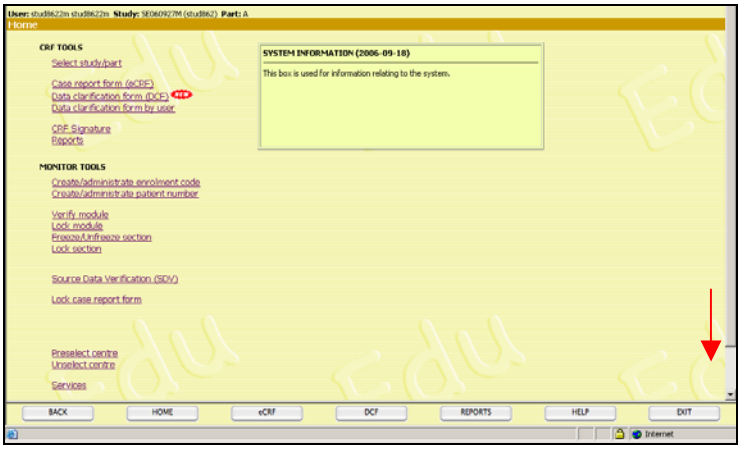

Figure 5.5 - Monitor start page

Consistency

The interface should have a consistent functionality and look throughout to make it easier to learn and use. Tables should be similar and the feedback that is received should be the same. An example is that the eCRF entry table looks very different compared to the SDV table.

These components need to be improved and used consistently throughout the system:

1. Button labels – The same terminology must be used when referring to the same function (submit for example must be used consistently)
2. Tables –The tables should look the same and operate using the same system as far as possible.
3. Color –Colors should be used with meaning and consistently throughout the interface.
4. Feedback – Feedback needs to follow one convention throughout the system to help the user get closure of activities and understand the system state.
5. The same structure should be used for different functions and sublevels. This is very important if the users are to be able to find their way through the system.

Alignment and Mapping

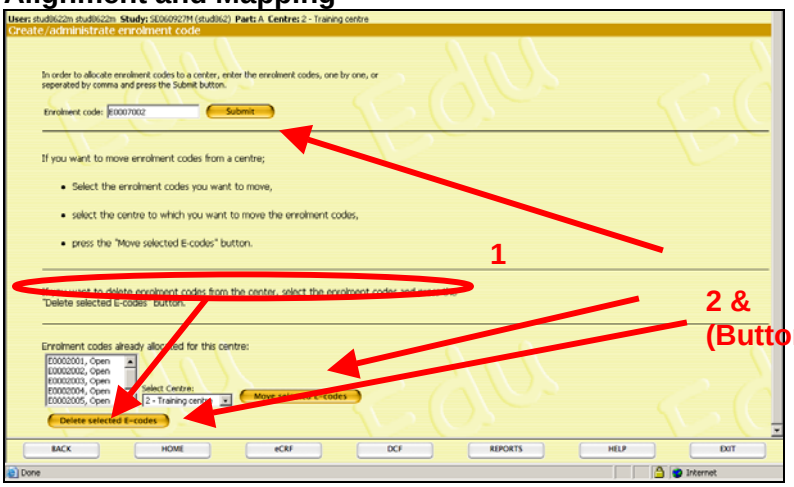

Figure 5.6 - Create administrate enrolment code

1. There is no natural mapping between descriptions and their function on several instances of the interface.
2. There is no use of a grid. The buttons should be aligned to create a sense that they all yield closure to the action taken.

Buttons

1. Inconsistent use of button labels. An example of this is the buttons in *figure 5.6 - Create administrate enrolment code.* Two of the buttons have a description as a label while the label on the third states “submit”. The labels should relate to their function and be consistent throughout the interface

Interaction with other systems

1. Study setup –
    - The study setup is partially conducted in the old paper process that was used before the introduction of this system. Support for the whole process in one system would make it more effective and simple.
    - Since it is very hard to make interface changes after study setup it is important that the interface design is well thought through before the study is launched.
2. StudyAdmin – The monitoring process uses both StudyCapture and the StudyAdmin system. It has been expressed that two computers are needed to run the systems in parallel. Being able to use both systems on one machine would help. Since a lot of data has to be used in both systems, better support for importing and exporting data would also help speed up and simplify the process.

Data checks

There are two main types of data checks. The types are, point-of-entry data checks and post-entry data checks. Several stakeholders have expressed that if there were more incorporated point-of-entry checks the process would be easier to handle. At the moment some of the data that has not been entered correctly is not detected until the post-entry data checks have been made overnight. After the overnight check the errors have to be corrected and checked again, which causes considerable delay in the process and extra work for the monitors and investigators.

Data corrections
In some cases it can be very hard and time-consuming to change faulty data. After a certain point in the data capture process the changes have to be administrated by the IS organization or by going through a special procedure were the responsible investigator makes the changes. This can be prevented by better data checks.

5.7.2 Key task usability problems

Log-on
Log-on problems are the most frequently reported problems according to the technical support department. Since some users do not use the system very often they often forget their log-on codes and encounter problems with their security solution (security card).

Data entry (investigator)
The data entry process is the most common process used by the investigators. Every subject is entered into the system. The data is entered into several parts. Problems in this process cause more work for the monitors who check the data as well as for the investigators who need to make the corrections.

eCRF (investigator)

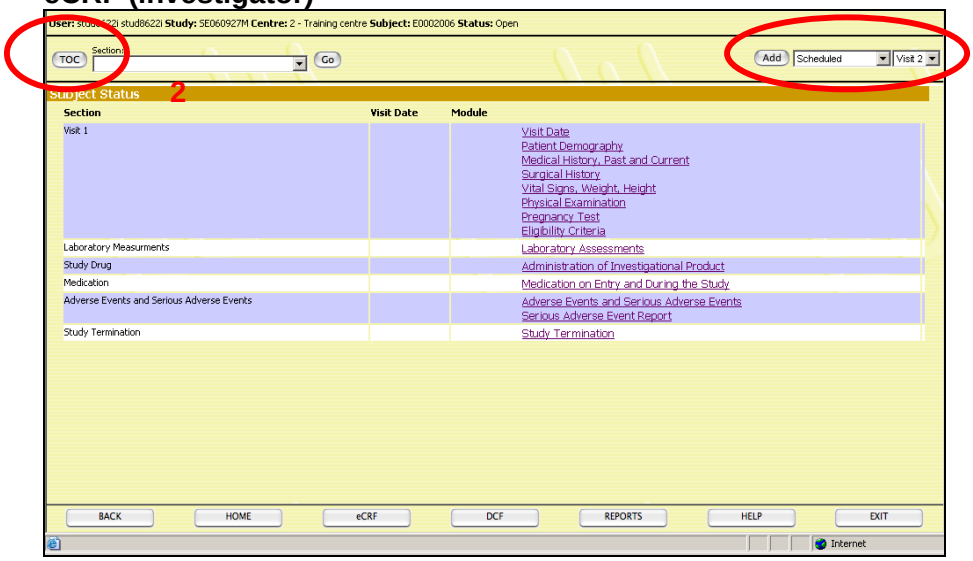

Figure 5.6 - ECRF (investigator)

1. The ”add” function is not described properly, what does it add? A label is missing. There is no natural chain of events either since the button "add" comes before the fields you are supposed to make choices in. The button should follow the choices.
2. The TOC button does not describe that it is TOC of the eCRF that it leads to. The user could be confused by the lack of structure and orientation help. Some indication of were in the structure the user came from and is presently would help.

Surgical History

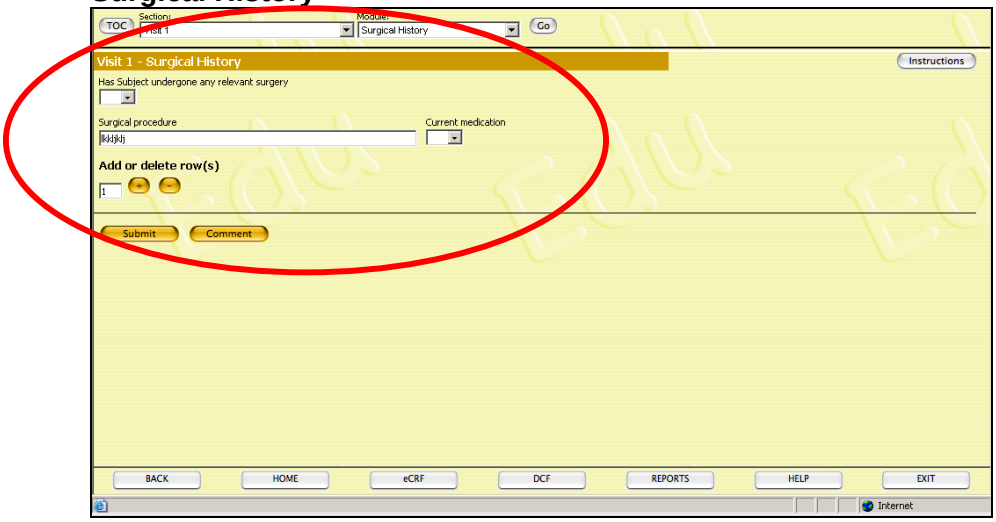

Figure 5.7 - Surgical History

- If "Has Subject undergone any relevant surgery" is left empty or "no" is chosen the user should not be able to enter data into the following fields to avoid mistakes. The fields that are not applicable should be of limits and indicted that they are inactivated.

Pregnancy Test

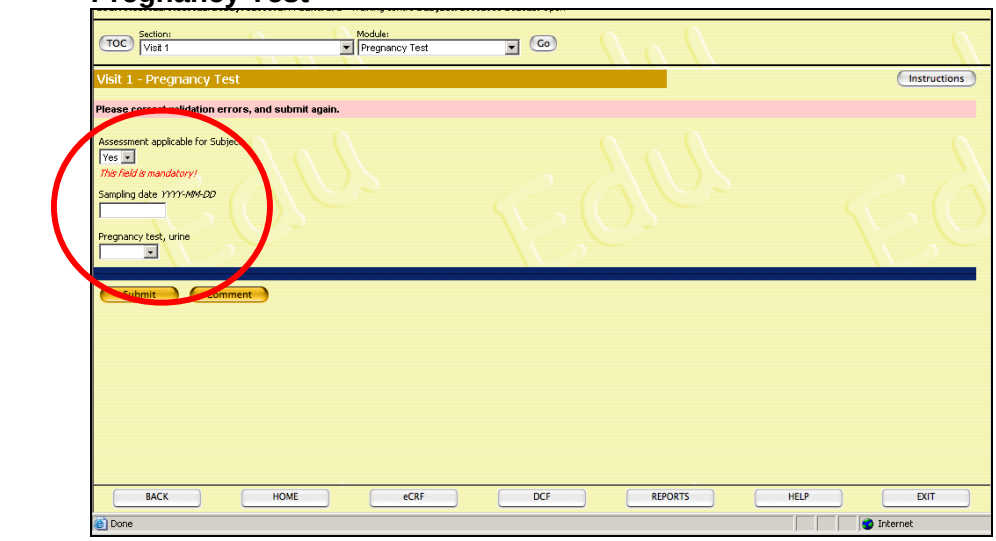

Figure 5.8 - Pregnancy Test

- If the subject has been entered as a male the pregnancy test should automatically be entered as not applicable. Not entering information yields an error in the checks, which has to be corrected later.

Medication

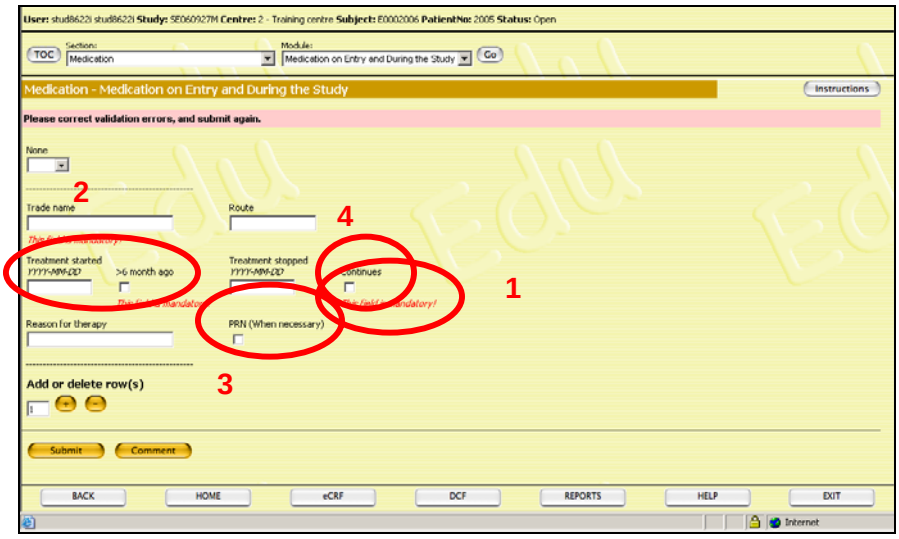

Figure 5.9 - Medication

1. The error messages should be descriptive and contain the information that the user needs.
2. Emphasize that it is either “Treatment started” date or “>6 months ago” that is to be filled in. It would be best if choices were grayed out if not applicable. If only one is to be entered this should be better indicted.
3. What does PRN stand for, write out abbreviations if they are not obvious. The PRN field should only be filled out if the other fields next to it are not filled in; this is not obvious in the interface.
4. Field labels – The labels on the fields do not explain their functions. The label should clearly indicate the meaning of the field. E.g. the "continues" field should be specified as ”continues after study termination” instead of just “continues”. According to two monitors users fill in this field in the wrong way 4/5 times.

Data checks (monitor)

Reports

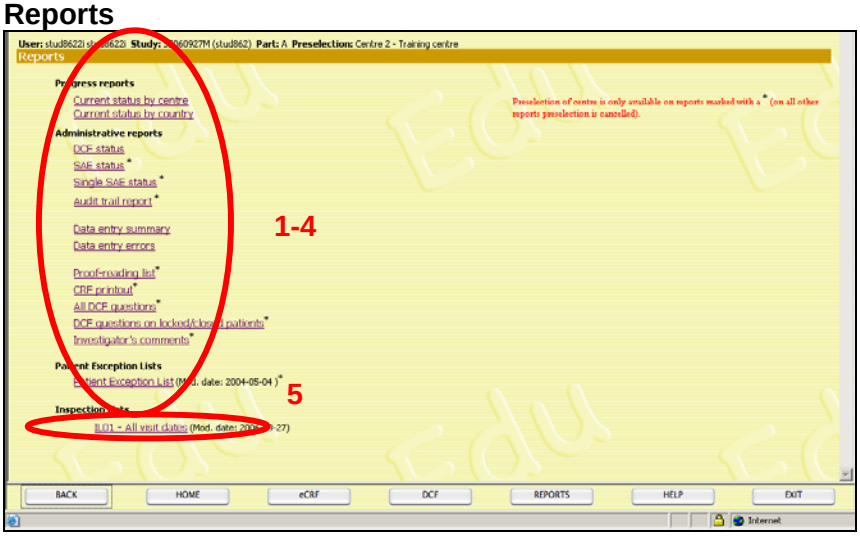

Figure 5.10 - Reports

1. Difficult for novice users to understand what abbreviations stand for and what the different reports stand for. Abbreviations and headlines that are not self

explanatory should be more specified and abbreviations should be changed into full words
2. The structure is not easily understood. There are too many links and their meaning and grouping is not clear enough. It would be better if there were fewer links on one page or if the links were better grouped.
3. The division into all types of different reports is quite confusing if you are not an expert user. The links do not follow a logical sequence based on their use.
4. The overview does not provide a sense of closure or feedback on how many errors that are left.
5. A link is hidden and you have to scroll down to see it, this link is easily missed.

Data entry errors

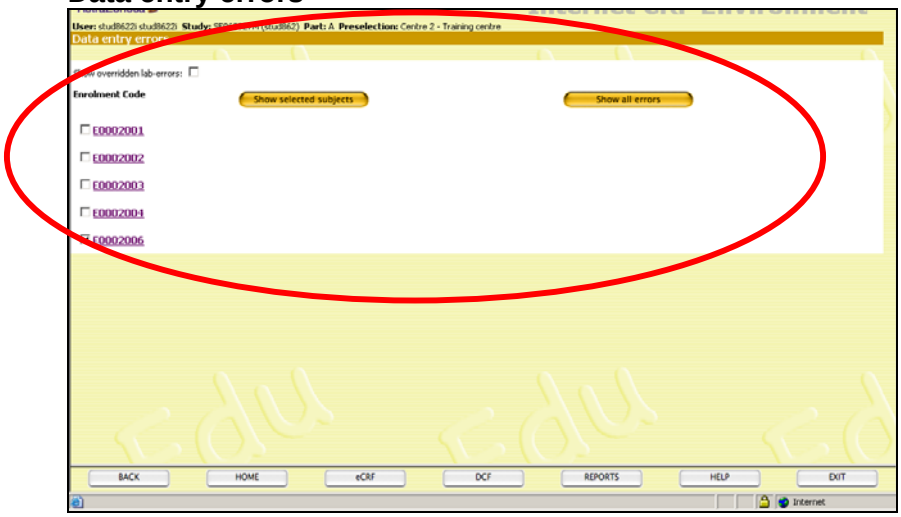

Figure 5.11 -Data entry errors

1. Once again the sequence is wrong between the buttons and the choices that are to be made. The tick boxes should come before the show-selected button since they should be ticked first. This problem creates confusion and does not give the user a sense of direction and closure.
2. Unnecessary mixture of white and yellow in the table leaves a fussy impression.

Patient exception list

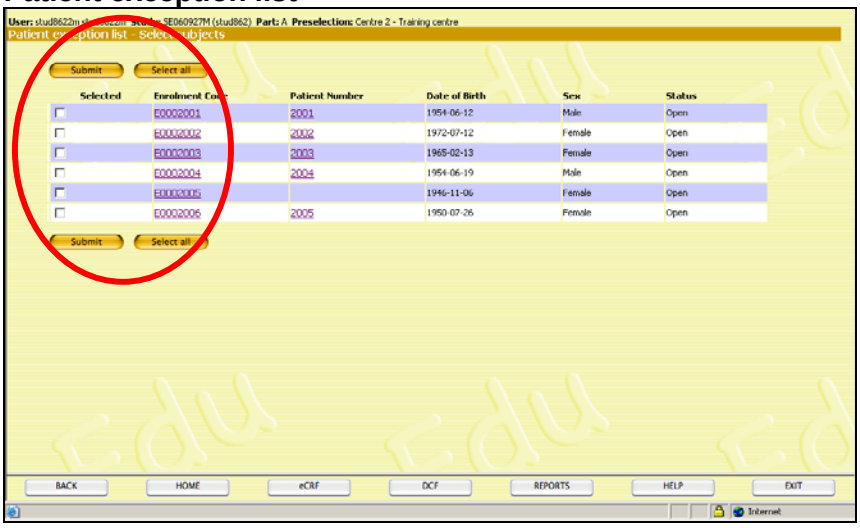

Figure 5.12 - Patient exception list 1

- Unnecessary use of two pairs of buttons, the buttons should follow the tick boxes to support the natural flow of actions.

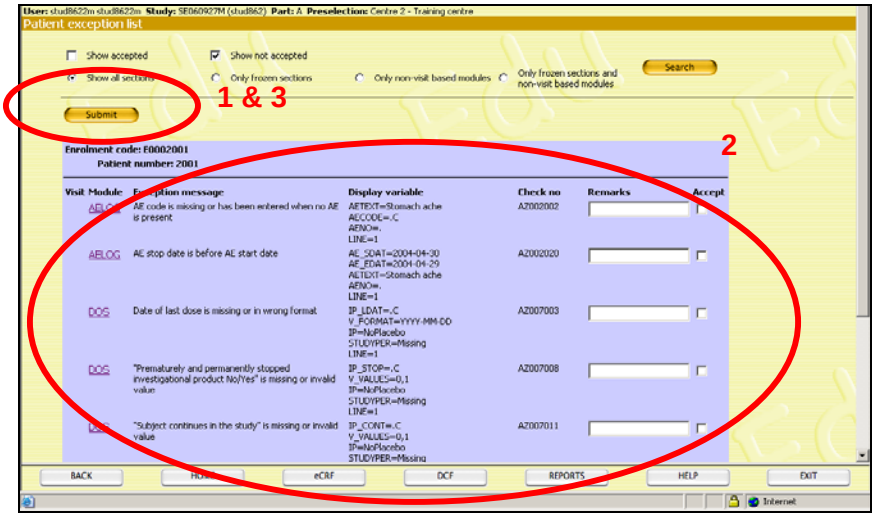

Figure 5.13 - Patient exception list 2

1. Again unnecessary submit button before the choices. Submit should give closure to the tasks consistently so that you never miss out on it and always know what actions have to be performed first.
2. Black/purple – blue contrast. The contrast is not good enough and makes the page less readable. The coloring is not consistent with other parts of the interface either.
3. There should be a description to emphasize that submit has to be made for the updates to be saved.
4. There is no feedback if the changes have been saved. This may cause the loss of data.
5. It is very complicated for monitors to translate the exceptions in the list into understandable questions to the investigators. The exception messages could be made more understandable and descriptive.

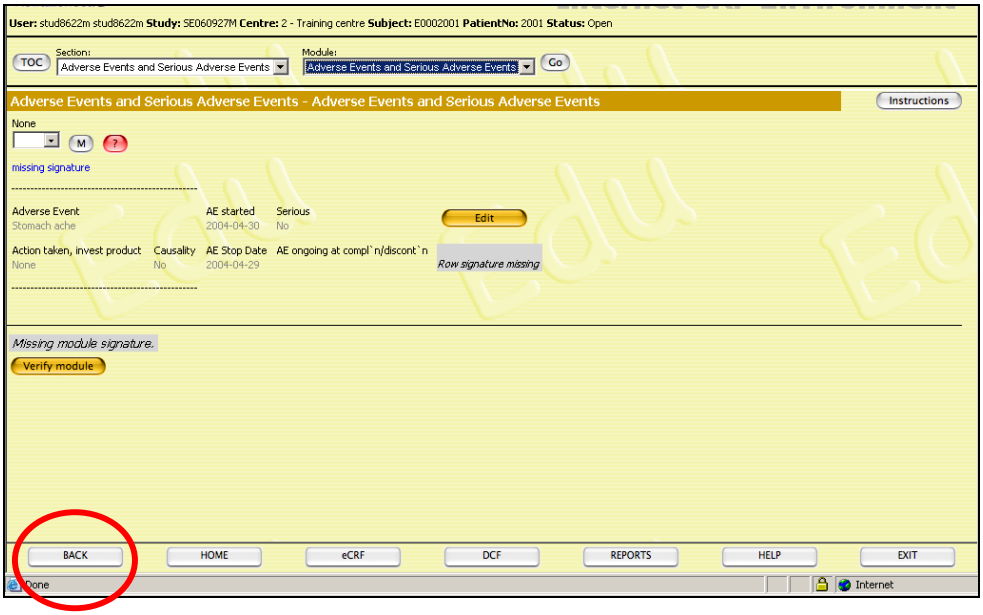

Figure 5.14 - Patient exception list 3

- When a patient exception is viewed it is hard for the user to find its way back to the overview of exceptions. There is no natural hierarchy of windows. The user is forced to use the back function. There needs to be a natural way of returning to the overview. This is another symptom of the structure problems of the interface.

SDV

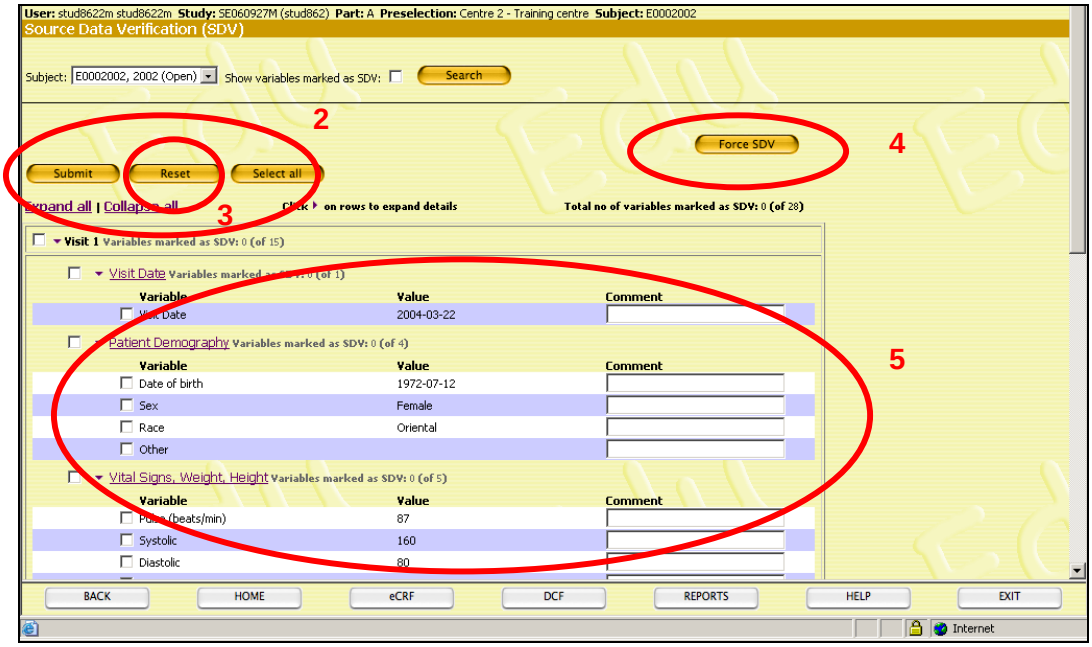

Figure 5.15 – SDV 1

1. When the user answers a query the information disappears and is not available in the query window. There should be a specification of the error still there while the query is being raised so that the user is reminded of what to write and what the error message was.
2. The placement of the buttons are confusing and there is a risk of data loss and re-entry of data if it is not clear when a submit must happen or if there is no feedback on what has been stored in the system.

3. The label on the “reset” button does not clearly describe what is reset when it is clicked.
4. There is no description of what “Force SDV” means.
5. There is a consistency problem throughout the system interface since many tables have a different look and functionality. This function for example could be built on the principles of the eCRF. In that interface the user even gets feedback on the status of the data through color-coding. Why not use the eCRF for verification why add another interface?

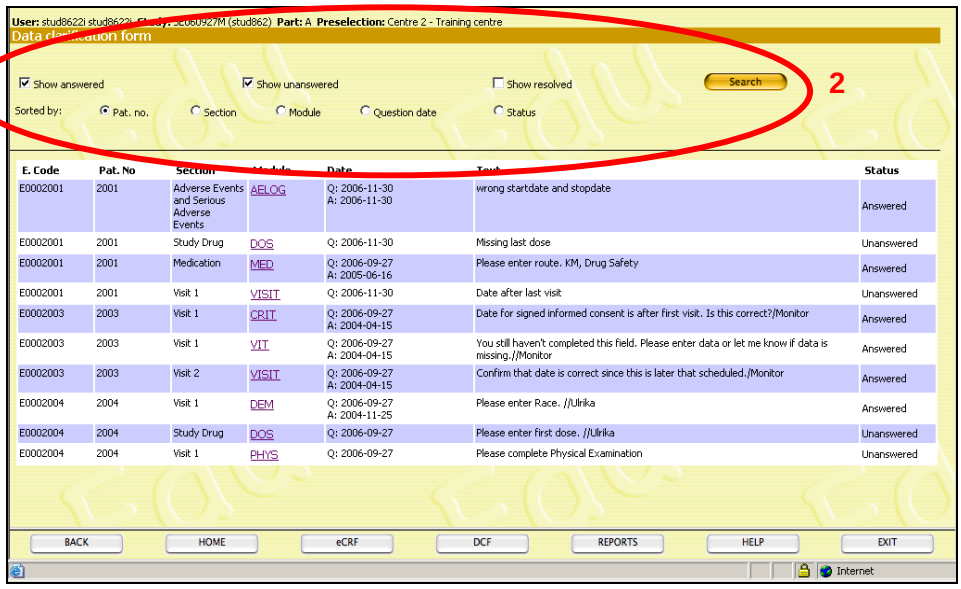

Figure 5.16 - SDV 2

1. Still the same problem with the structure, how does the user get back to the overview after entering a specific problem?
2. Grouping and alignment of search alternatives might confuse the user. The layout should be more thought through and use a grid to get natural grouping. It is not clear either that the search button is connected to the alternatives; this could also be corrected by grouping the alternatives better.

.

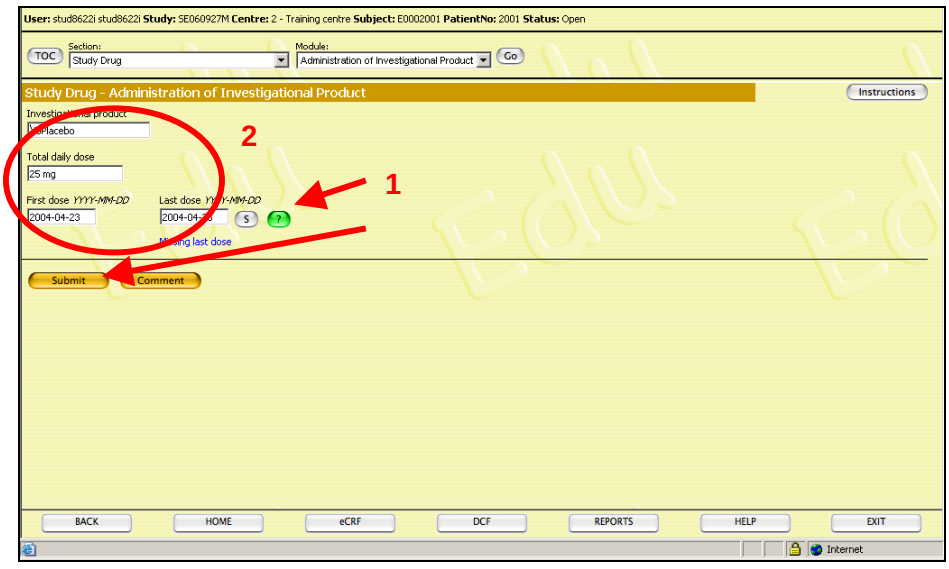

Figure 5.17 - SDV 3

1. There is not a simple closure to the dialog. The user has no way of knowing if the query should be answered first or if the value change should come first. Submit should be a closure to the action.
2. Feedback when things have been changed needs to be better. Text that needs to be submitted needs to stand out form submitted text. This would help the users know when to submit and thereby save time and data loss.
3. The sequence of action should be indicated. Perhaps by only allowing the value change after a query is answer.

SAE Handling

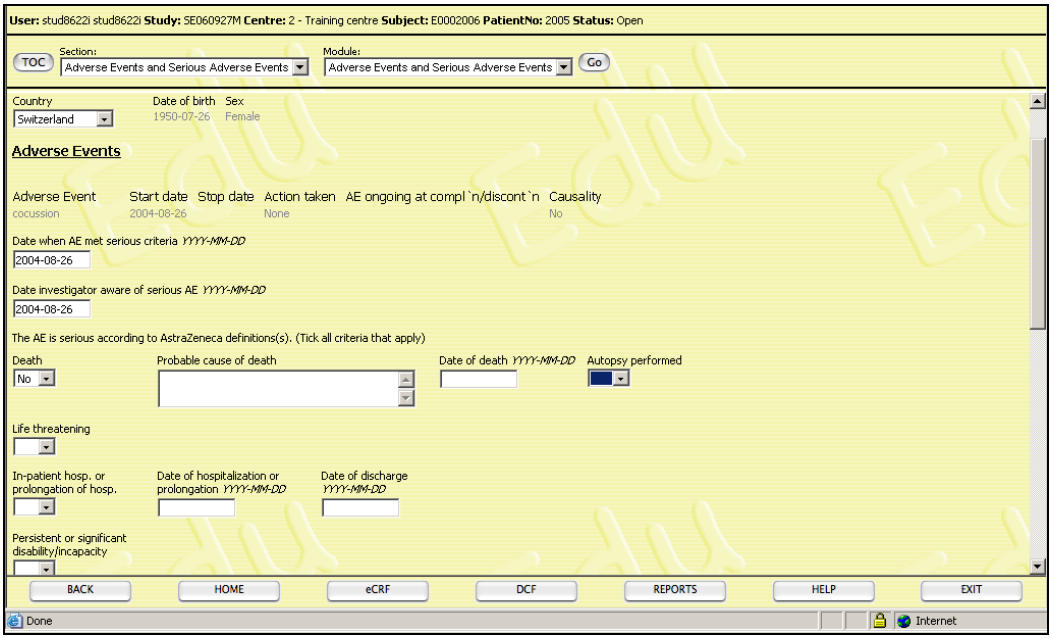

Figure 5.18 - Adverse events and serious adverse events

- Lock fields that are not supposed to be used if another field is filled in.

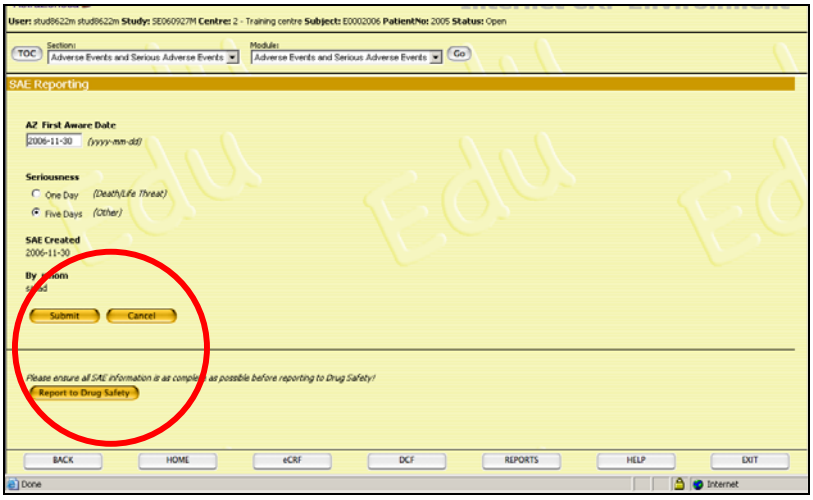

Figure 5.19 - Monitor SAE

- The user has to re-enter the SAE after submitting and then report to drug safety. There is no closure and clear chain of events; it is therefore easy to miss one or the other.

5.8 Usability problem report 2 – With severity ratings

Help Functions

Usability problem 1 - Help

Frequency of use	1.5
Error-frequency	2
Impact	1
Severity rating	3

If the user needs help with a specific part of the process he or she needs to find it in the help section. This makes it much harder and more time-consuming to find than if the help section is connected to the screen that is viewed at the time help is requested.

Study set-up

Frequency of use	0.5
Error-frequency	2.5
Impact	1.5
Severity rating	1.875

Usability problem 2 - Integrated set-up

The study-setup is partially conducted in the old paper process that was used before the introduction of this system. Support for the whole process in one system would make it more effective and time could be saved.

Feedback, Orientation and structure

3

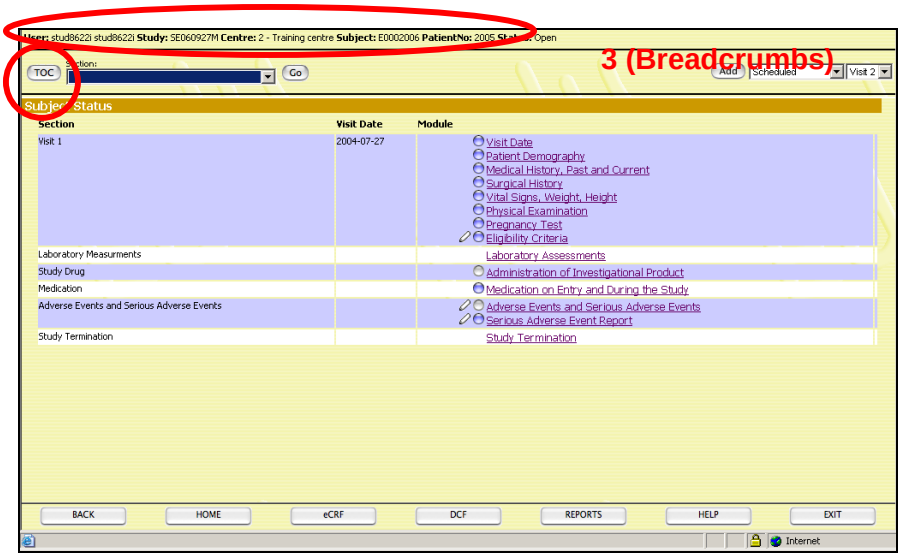

Figure 5.20 – Feedback, Orientation and structure example – eCRF screen

Usability problem 3 - Breadcrumbs and levels

Frequency of use	3
Error-frequency	3
Impact	2
Severity rating	18

The system is using breadcrumbs (3) to help system orientation. However they are not used consistently. This functionality can be improved by using breadcrumbs for all levels and to use them consistently. If they were made into links it would help the user get a sense of location and orientation throughout the system. In combination with the fact that there are different ways to reach the same system states, this may be confusing for the user. The structure of the system and levels and sublevels needs to be evident. The users always have to know were in the structure they are.

- An example of a problem that could be avoided with better breadcrumb use is the *Page labels* **-** There should not be a label called “home” or a label called “TOC” links should refer to the page by name. It is not clear which page TOC refers to.
- There is no distinction between system levels. It makes it hard to distinguish between entry and overview pages. Use colors, or other distinctions to distinguish different levels and functions in the system. This could be done carefully and with minor variations of color, for example.

Alignment and Mapping

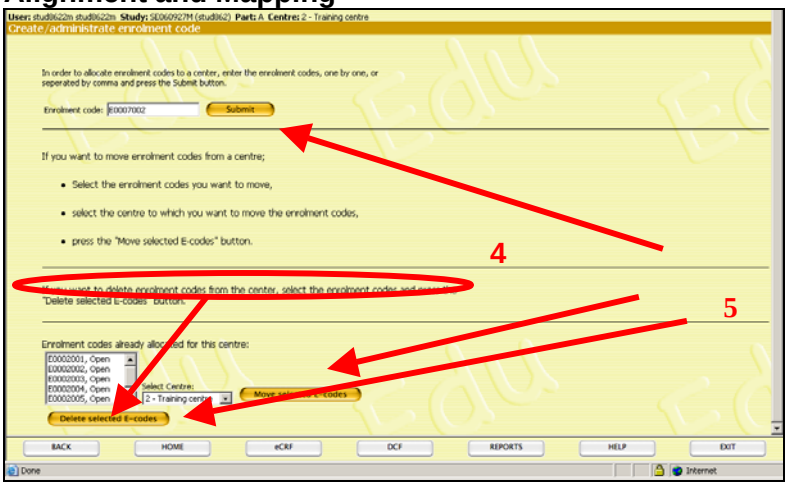

Figure 5.21 – Alignment and Mapping example - Create administrate enrolment code screen

Usability problem 4 – Descriptions

Frequency of use	2.5
Error-frequency	2
Impact	1
Severity rating	5

Throughout the interface the mapping between descriptions and their functions are flawed. The fact that the items are not grouped properly makes it harder for the user to understand how to use the interface. There is also a chance that the user misreads or misses the descriptions. There is no natural connection between descriptions and their function on several instances of the interface.

Usability problem 5 - Structure and sequence

Frequency of use	2.5
Error-frequency	2.5
Impact	1.5
Severity rating	9.375

Grids are not used consistently throughout the interface to improve readability by aligning items. For example buttons should be aligned and arranged to create a

sense that they all yield closure to the action taken. The right grouping and sequence of buttons can help the user understand what actions are expected and in what order. In many cases in the system buttons for submit; come before the choices that they submit. This makes it more likely that the users submit to early or forget to submit the data.

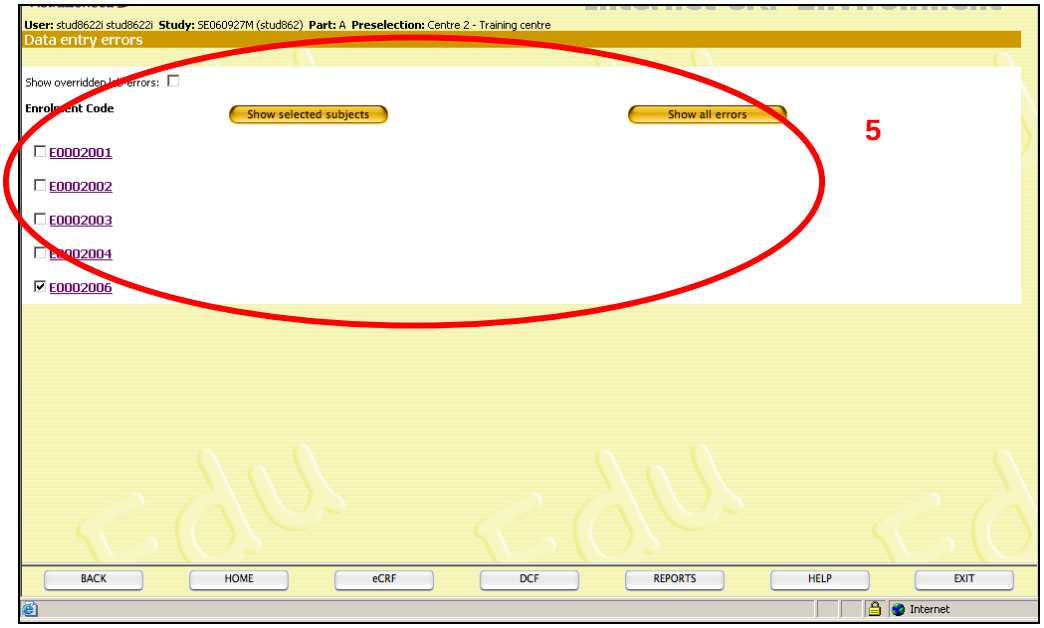

Figure 5.22 - Structure example. -Data entry errors screen

Terminology

Usability problem 6 - Abbreviations and terminology

Frequency of use	2.5
Error-frequency	2
Impact	1
Severity rating	5

It is not recommended that abbreviations be used in an interface unless they are absolute convention. In many cases throughout the system the meaning of the abbreviations might not be clear to all users. Writing out abbreviation puts less strain on the users memory. There also needs to be a list of accepted abbreviations for input that users can refer to and find easily. All text in the system should be adapted to fit the users; terminology that is familiar to them should be used.

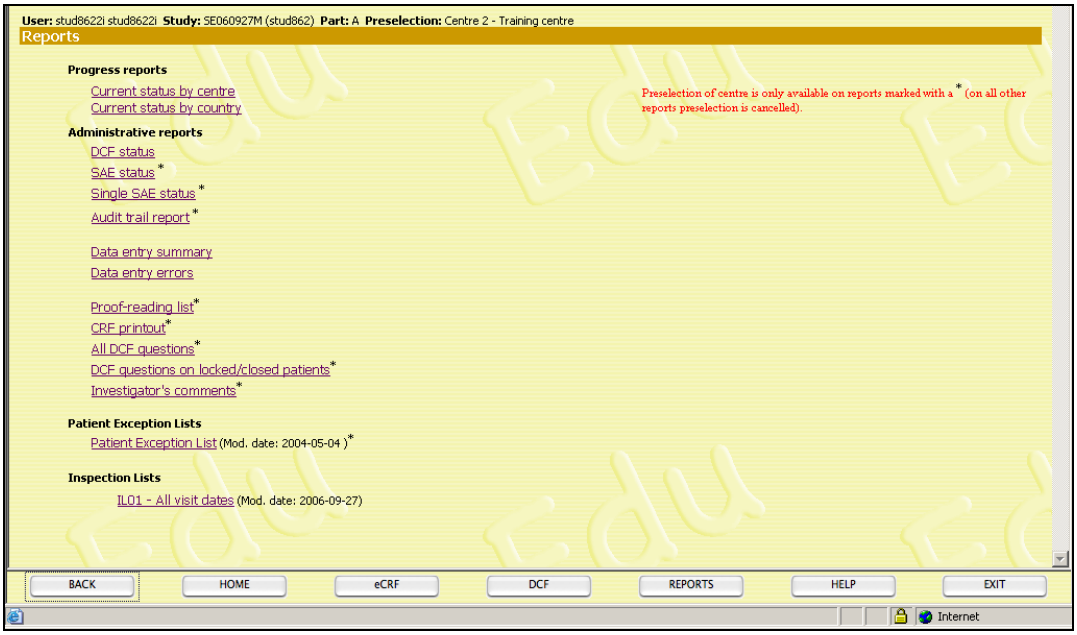

Figure 5.23 - Terminology example: Many abbreviations – reports screen

Usability problem 7 - Consistency

Frequency of use	2.5
Error-frequency	2
Impact	1
Severity rating	5

The interface should have a consistent functionality and look throughout, to make it easier to learn and use. If similar components are used in different ways it makes it extremely difficult for the user to learn and remember how to use the system. Tables should be similar and the feedback that is received should be the same. An example is that the eCRF entry table looks very different compared to the SDV table.

These components have consistency flaws throughout the system:

- Button labels – The same terminology must be used when referring to the same function (“submit”, for example is not used consistently throughout the system)
- Tables –The tables should look the same and operate using the same system as far as possible.
- Color – Colors should be used with meaning and consistently throughout the interface.
- Feedback – Feedback needs to follow one convention throughout the system to help the user get closure of activities and understand the system state.
- The same structure should be used for similar functions and sublevels. This is very important if the users are to be able to find their way through the system.

Messages & Text

Usability problem 8 - Messages

Frequency of use	2.5
Error-frequency	2
Impact	1
Severity rating	5

- Visibility - Italic text in the interface is not emphasized enough, in areas where several choices are possible an italic text specifies if several alternatives are to be chosen or if the user should pick only one. In that scenario the possibilities are indicated by an “or” or an “and” between the choices. If the text is not emphasized better the user might mark to few or to many alternatives, which will result in extra time to correct errors
- Status messages -There is no distinction between error messages and status messages. The use of feedback text in red is questionable when an action was successful. Convention is that red indicates failure. Also the feedback text is not grouped with the action taken which can make it difficult for the user to understand what caused the problem.

Data checks

Usability problem 9 - Data checks

Frequency of use	2.5
Error-frequency	3.5
Impact	2
Severity rating	17.5

There are two main types of data checks. The types are, point-of-entry data checks and post-entry data checks. Several stakeholders have expressed that if there were more incorporated point-of-entry checks the process would be easier to handle. At the moment some of the data that has not been entered correctly is not detected until the post-entry data checks have been made overnight. After the overnight check the errors have to be corrected and checked again, which causes considerable delay in the process and extra work for the monitors and investigators. By extending the point-of-entry data checks the time spent asking and answering questions can be greatly reduced for both monitors and investigators.

Usability problem 10 - Log-on

Frequency of use	3.5
Error-frequency	2
Impact	2
Severity rating	14

Log-on problems are the most frequently reported problems according to the technical support department. Since some users do not use the system very often they often forget their log-on codes and encounter problems with their security solution (security card). Changing the logon to a different method could reduce time spent logging on for users as well a save time for the support staff.

Usability problem 11 – Feedback

Frequency of use	3.5
Error-frequency	3
Impact	2
Severity rating	21

It is hard to distinguish data that has been saved and submitted from data that has been inserted but not submitted. The fact that this problem occurs throughout the system raises the risk for a loss of data.

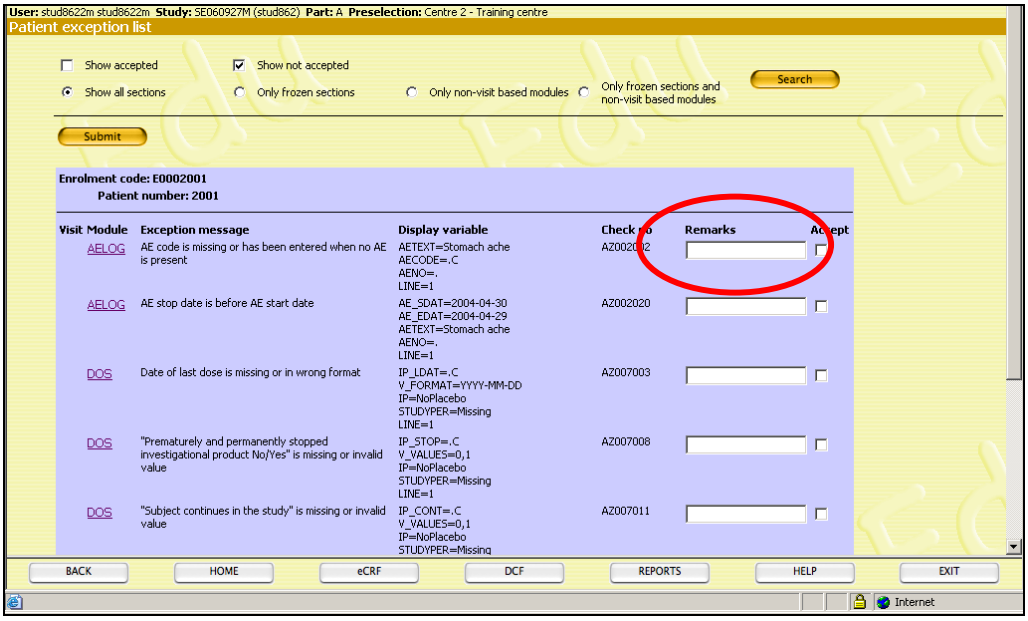

Figure 5.24 - Feedback example – If information is entered into the field it will be present next time the page is accessed. If additional data is entered then there will not be any way of knowing what has been submitted.

5.9 Usability problem report 3 – With CBA

This report contains usability problems that have been found in the StudyCapture system. The problems in this report are the problems that were judged as the most severe based on an evaluation and severity rating of the system.

The following five problems were selected for a CBA since they received the highest severity rating.

Table 5.8 – Ranked severity rating

Problem 11	21
Problem 3	18
Problem 9	17.5
Problem 10	14
Problem 5	9.4

The usability problems are listed and described below with the severity ratings that they have received and the sub-factors that go into the severity rating calculations. In addition to the rating the results of the CBA meeting are presented. For each problem the benefit market is calculated in FTEs based on an analysis of the consequences of the problem. The calculations for each problem are based on the average time that the CBA meeting participants estimated that each consequence would consume as well as for the minimum time that was also estimated for each consequence. In the illustrations the minimum time is indicated within parenthesis.

Usability problem 11 – Feedback

Problem description:

It is hard to distinguish data that has been saved and submitted from data that has been inserted but not submitted. The fact that this problem occurs throughout the system raises the risk for a loss of data.

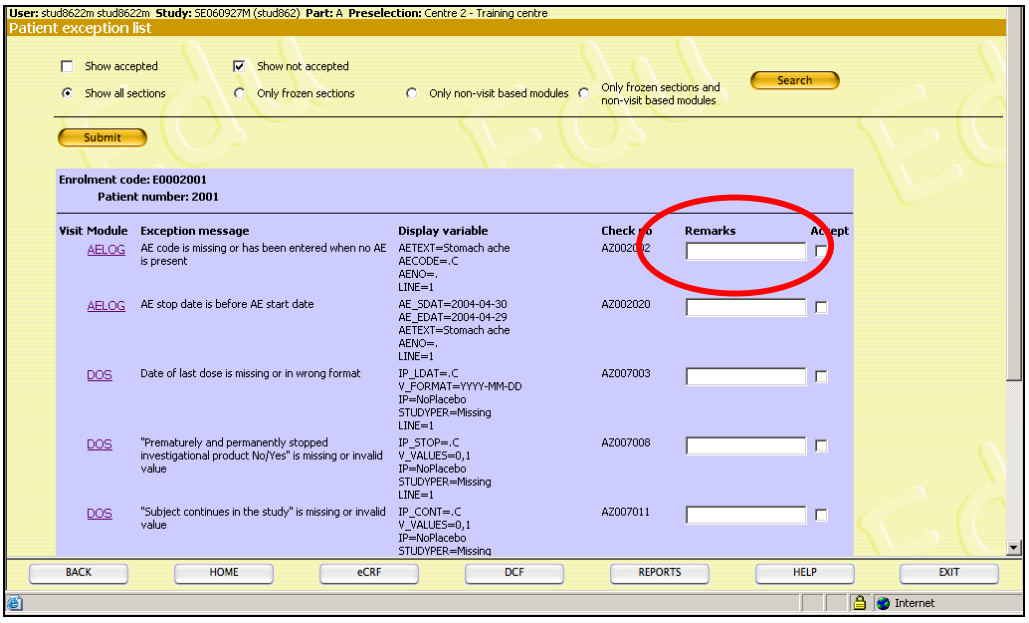

Figure 5.25 - Feedback example – If information is entered into the field it will be present next time the page is accessed. If additional data is entered then there will not be any way of knowing what has been submitted.

The following table displays the results from the severity rating activity:

Frequency of use	3.5 (daily - several times a day)
Error-frequency	3 (often)
Impact	2 (serious)
Severity rating	21

Based on the severity rating and the discussion at the CBA meeting the following CBA was made:

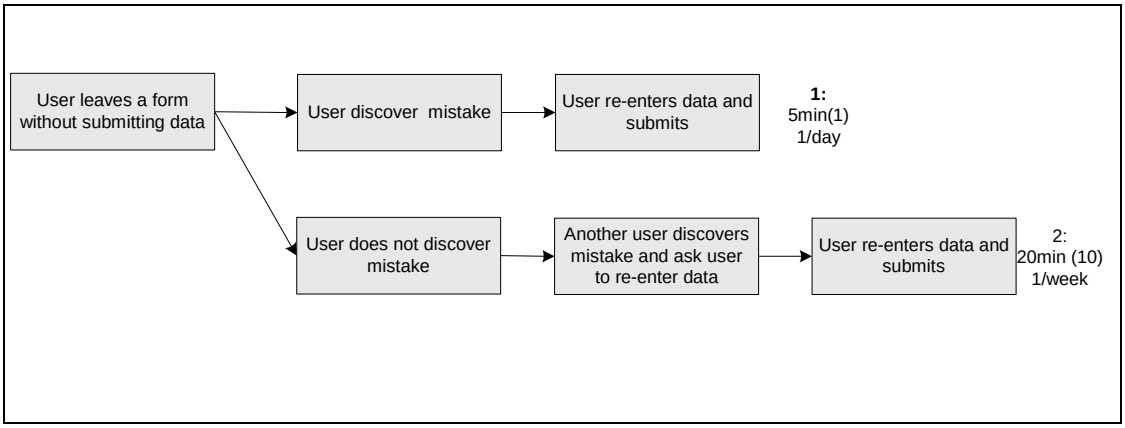

Figure 5.26 - Analysis of the problem consequences

Table 5.9 - Benefit market for the problem:

	Average FTE cost	Minimum FTE cost
Consequence 1	8	1.6
Consequence 2	6.4	3.2
Total	14.4	4.8

Usability problem 3 - Feedback, Orientation and structure
Problem description:

3

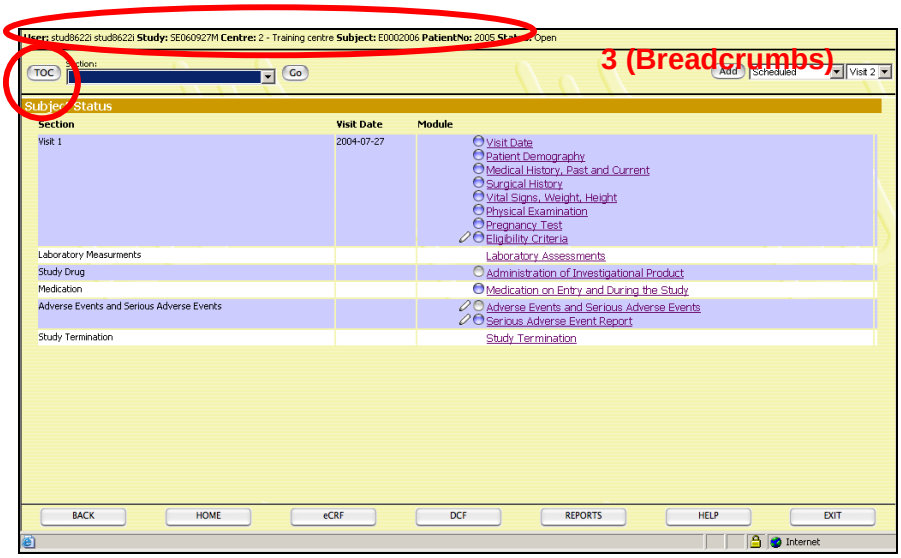

Figure 5.27 - Feedback, Orientation and structure example – eCRF screen

Breadcrumbs and levels - The system is using breadcrumbs (3) to help system orientation. However they are not used consistently. This functionality can be improved by using breadcrumbs for all levels and to use them consistently. If they

were made into links it would help the user get a sense of location and orientation throughout the system. In combination with the fact that there are different ways to reach the same system states, this may be confusing for the user. The structure of the system and levels and sublevels needs to be evident. The users always have to know were in the structure they are.

- An example of a problem that could be avoided with better breadcrumb use is the **Page labels -** There should not be a label called “home” or a label called “TOC” links should refer to the page by name. It is not clear which page TOC refers to.
- There is no distinction between system levels. It makes it hard to distinguish between entry and overview pages. Use colors, or other distinctions to distinguish different levels and functions in the system. This could be done carefully and with minor variations of color, for example.

The following table displays the results from the severity rating activity:

Frequency of use	3 (daily)
Error-frequency	3 (often)
Impact	2 (serious)
Severity rating	18

Based on the severity rating and the discussion at the CBA meeting the following CBA was made:

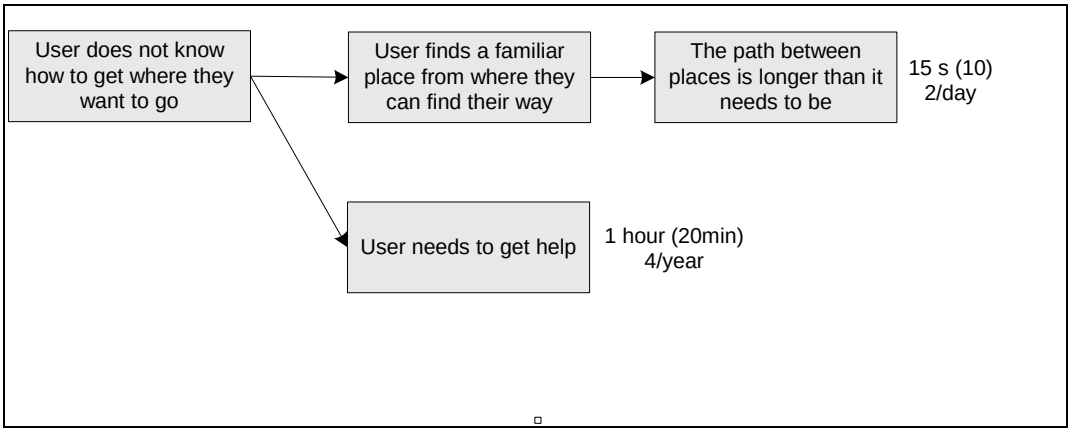

Figure 5.28 - Analysis of the problem consequences

Table 5.10 - Benefit market for the problem:

	Average FTE cost	Minimum FTE cost
Consequence 1	0.8	0.54
Consequence 2	0.48	0.16
Total	1.28	0.7

Usability problem 9 - Data checks

Problem description:

Data checks – There are two main types of data checks. The types are, point-of-entry data checks and post-entry data checks. Several stakeholders have expressed that if there were more incorporated point-of-entry checks the process would be easier to handle. At the moment some of the data that has not been entered correctly is not detected until the post-entry data checks have been made overnight. After the overnight check the errors have to be corrected and checked again, which causes

considerable delay in the process and extra work for the monitors and investigators. By extending the point-of-entry data checks the time spent asking and answering questions can be greatly reduced for both monitors and investigators.

The following table displays the results from the severity rating activity:

Frequency of use	2.5 (weekly - daily)
Error-frequency	3 (often - every time
Impact	2 (serious)
Severity rating	17.5

Based on the severity rating and the discussion at the CBA meeting the following CBA was made:

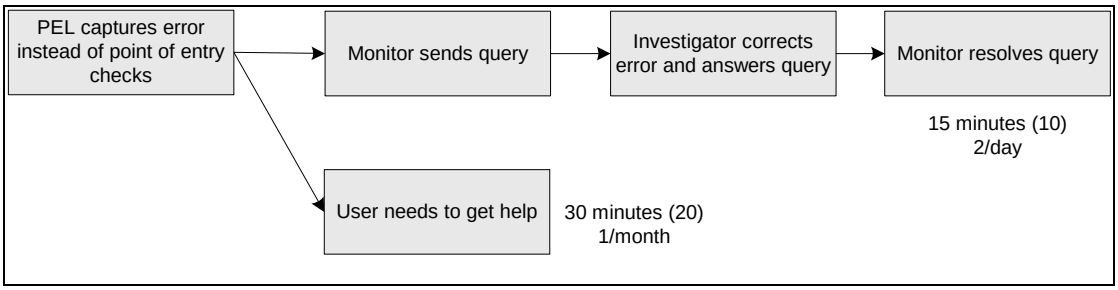

Figure 5.29 - Analysis of the problem consequences

Table 5.11 - Benefit market for the problem

	Average FTE cost	Minimum FTE cost
Consequence 1	48	32
Consequence 2	2.2	1.5
Total	50.2	33.5

Usability problem 10 - Log-on
Problem description:
Log-on problems are the most frequently reported problems according to the technical support department. Since some users do not use the system very often they often forget their log-on codes and encounter problems with their security solution (security card). Changing the logon to a different method could reduce time spent logging on for users as well a save time for the support staff.

The following table displays the results from the severity rating activity:

Frequency of use	3.5 (daily - several times a day)
Error-frequency	2 (every now and then)
Impact	2 (serious)
Severity rating	14

Based on the severity rating and the discussion at the CBA meeting it was decided that the benefit market was not to be calculated for this problem since it was not a usability problem.

Usability problem 5 - Structure and sequence
Problem description:
Grids are not used consistently throughout the interface to improve readability by aligning items. For example buttons should be aligned and arranged to create a

sense that they all yield closure to the action taken. The right grouping and sequence of buttons can help the user understand what actions are expected and in what order. In many cases in the system buttons for submit; come before the choices that they submit. This makes it more likely that the users submit to early or forget to submit the data.

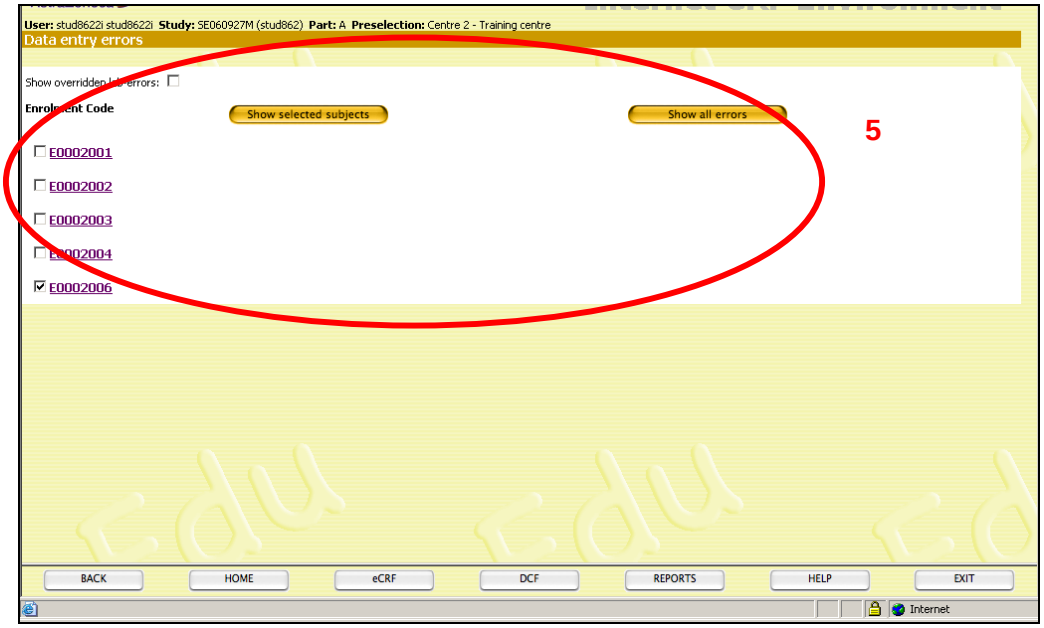

Figure 5.31 - Structure example. - Data entry errors screen

The following table displays the results from the severity rating activity:

Frequency of use	2.5 (weekly - daily)
Error-frequency	2.5 (every now and then - often)
Impact	1.5 (minor - serious)
Severity rating	9.4

Based on the severity rating and the discussion at the CBA meeting the following CBA was made:

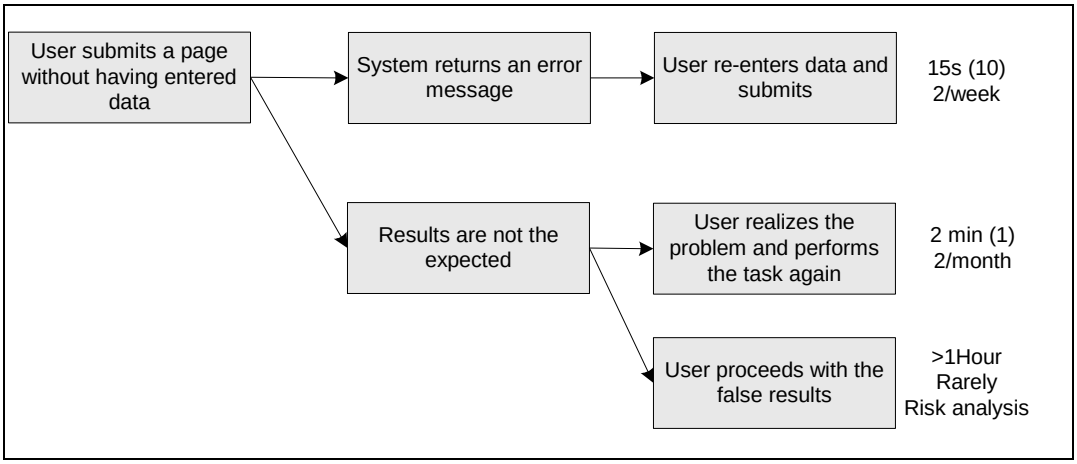

Figure 5.32 - Analysis of the problem consequences

Table 5.12 - Benefit market for the problem

	Average FTE cost	Minimum FTE cost
Consequence 1	0.16	0.11
Consequence 2	0.30	0.15
Consequence 3	?	?
Total	0.46	0.15

6 Case study – StudyAdmin

A case study for the StudyAdmin system was conducted to test the first framework proposal. The case study gives the evaluator a chance to see how the framework works in practice. A secondary goal was to get a usability evaluation of the StudyAdmin system, complete with a cost-benefit analysis of the most severe usability problems that were found.

6.1 System description

StudyAdmin is a Study Management tool that is used globally within AstraZeneca R&D. The system is externally developed and is used by eleven other major pharmaceutical companies as well. Mainly monitors and study management teams use the StudyAdmin system. It provides an overview of the study's progress and if deadlines are met. StudyAdmin is mainly used to perform source data verifications to help AZ staff to follow the patients that are in the medical study.

6.2 Background studies

A thorough background study was conducted in order to give the evaluator broader domain knowledge. This is necessary to have before the evaluator can derive good test cases.

The main goals of the background studies phase were that the evaluator was to acquire an understanding of StudyAdmin and the roles of the staff surrounding it. A secondary goal was for the evaluator to form an opinion of the objectives that the system was supposed to fulfil, as well as the most critical user-tasks that the system was supposed to support.

6.2.1 Start-up

A first meeting with AZ staff took place where it was decided that StudyAdmin was to be the object for evaluation. A short brain storming session resulted in a number of suggested activities and a number of people to contact, so that the evaluator could gain a better understanding of StudyAdmin. A number of questions were also raised, such as:

- How many people do you need to contact to say that you have done a sufficient background study?
- How much time do you need with the system, in order to get acquainted with it?
- What are the goals for the system?

After the consultancy a decision was made that a short description of the system was needed. This description was to be held by an administrator of the system. An interview with a so-called super-user was also needed to get more insight into the system. Finally a demonstration at a monitoring visit was to be conducted.

6.2.2 Description of StudyAdmin

An hour-long appointment with an AZ staff with enough domain knowledge was arranged in order for the evaluator to gain more of an overview of the system. The system architecture was described and the organizational structure was explained. Several other questions were answered. I.e.

- What types of different users are there?
- What kind of education do you need to operate StudyAdmin?
- What kind of documentation is there to study?

6.2.3 Interview with a super-user

To collect more specific and technical information regarding StudyAdmin an interview with a super-user was held. A super-user is a person with extensive domain knowledge. A super-user has a supporting role in the organization, supporting the monitors with technical problems. There were a long discussion about what the super-user thought were problems within StudyAdmin and she gave a couple of small hints on what to focus the evaluation on.

The interview led to that the evaluator gained more understanding to what the problems with StudyAdmin were, a long with a better understanding of the system itself and what things needed to be studied.

6.2.4 Monitoring visit

A monitoring visit was arranged in order for the evaluator to gain a more insightful view into the monitor's work process. A monitoring visit is a visit made by the monitor at a care centre. At the care centre the monitor performs source data verifications on the case report forms that the investigator has filled out. The case report forms is a report form where medical data about a patient, enrolled in a clinical study, is documented.

The purpose of a monitoring visit is to check that the investigators perform the correct procedures during the clinical trial.

The purpose for the evaluator to observe the visit was so that he could observe in detail what the monitor was working with. The workflows were studied and documented. These workflows later formed a basis for the heuristic evaluation.

There was also a chance to conduct an interview with the monitor. This gave the evaluator a chance to hear what an actual user thought of the system and where she had problems understanding it. A few of the problems presented by the monitor are in the problem report and were analyzed with extra care by the evaluator.

6.3 Evaluation

6.3.1 Attitude survey

An attitude survey was conducted during the evaluation phase. This survey gave the evaluator the chance to gain an overall attitude towards the usability in StudyAdmin. The attitude survey was taken from a template especially made for this purpose. The system usability scale (Brooke, J. 1986) (SUS) gives the evaluator composite measure of the overall usability for the system being studied.

A total of 15 surveys were collected and analyzed.
The distribution of the participants is as below:

- Eleven persons works with monitoring
- Eight persons works as a super-user
- Two persons works as a system administrator

Note that a person can have more than one role within the system. The result from the survey is presented below.

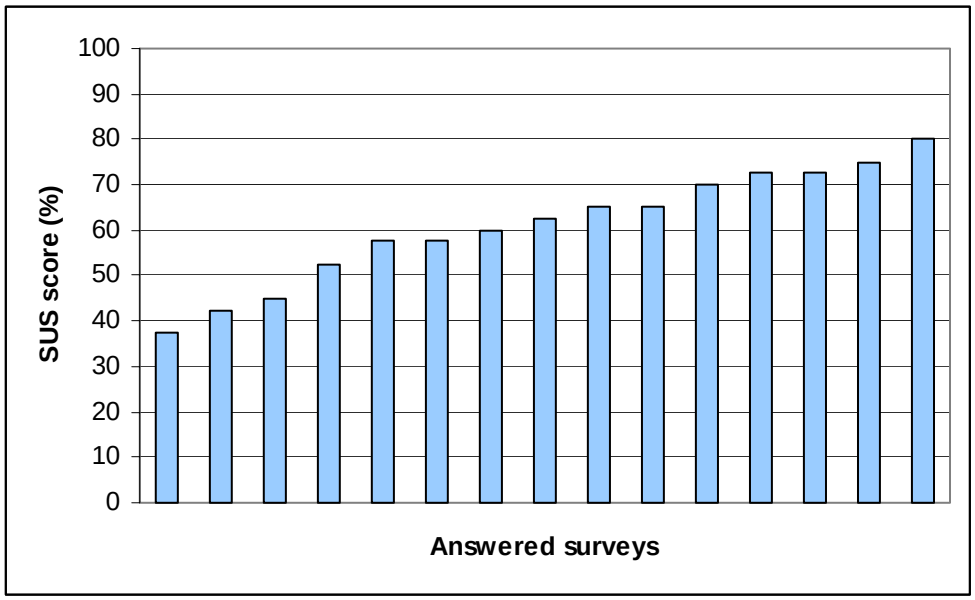

Figure 6.1 - SUS scores of all surveys

There was a slight difference between the different user groups in attitude towards the usability of the system.

Table 6.1 - Attitude survey responses

	Monitors:	Super-users:	Administrators:
Attitude mean:	62.5	60.6	56.3
Attitude median:	65	62.5	56.3

As the results above shows, it seems that the more technical knowledge you gain the less you think of the usability of the system.

Table 6.2 - Work experience

	Monitors:	Super-users:	Administrators:
Mean work exp.	3.5	3.6	2.5
Mean hours per week spent:	9	5.9	3

Table 6.2 shows the number of years the different types of users have worked on an average with the system. It also shows how many hours a week they spend.

Table 6.3 - Attitude vs. work experience

	Monitors:	Super-users:	Administrators
Attitude mean (less than 3.5 ys exp.):	63.3		
Attitude mean (less than 3.6 ys exp.):		57.5	
Attitude mean (less than 2.5 ys exp.):			37.5
Attitude mean (more than 3.5 ys exp.):	61.5		
Attitude mean (more than 3.6 ys exp.):		55.8	
Attitude mean (more than 2.5 ys exp.):			75

Table 6.3 shows that there is a slight difference in the attitude compared to how long they have worked with the system. The only exception is the administrators, but since there are only two answers by administrators, no real conclusions can be made from that result. This means that the longer a user have worked with the system the more he dislikes the usability. There can be different reasons for this. One can be that since they have worked for a long time with the system, they have also noticed competitors to the system in use. Or maybe they have worked with previous systems and found that they had a better usability.

The number for the number of years of work experience is chosen from the mean value for each user group.

Table 6.4 - Attitude vs. time spent per week

	Monitors:	Super-users:	Administrators:
Attitude mean (less than 9 hs spent):	57.1		
Attitude mean (less than 6 hs spent):		55	
Attitude mean (less than 6 hs spent):			37.5
Attitude mean (more than 9 hs spent):	60.8		
Attitude mean (more than 6 hs spent):		61	
Attitude mean (more than 6 hs spent):			75

Table 6.4 shows the attitude vs. the time spent per week. It clearly shows that the more time a user spend with the system the better attitude towards the system the user gets. This can be interpreted as that the user learns its ways around problems that arise and is therefore not bothered by them anymore. If a user does not spend as much time as needed to do this it will always face the same problems and will therefore dislike the usability more.

The number of hours spent is chosen from the mean number of hours spent for each user group.

6.3.2 Heuristic evaluation

The evaluator conducted a heuristic evaluation during the evaluation phase. This was done in order to get an experts view of the system and to collect the usability problem to work with during the cost-benefit analysis.

The workflow had been studied at the monitoring visit, where the system had been briefly explained and demonstrated. This workflow served as a basis for the deriving of test cases used in the evaluation. The heuristic evaluation took principles like Norman's *Design principles,* Nielsen's *Heuristic guidelines* and Shneiderman's *Eight golden rules of interface design* in considerations. These principles together with the use cases then formed the work process for the evaluator. The method chosen for

the evaluation became therefore a sort of cognitive walk-through where every error or problem was documented along the way.

When the evaluation came to an end a problem report was made. The problem report is placed under chapter *6.7 – Usability problem report 1 – First summary.*

6.4 Usability severity rating

To gain good enough validity, it is said that at least four persons has be a part of the severity rating process. Therefore four people, including the evaluator, were set to perform the severity ratings. To start this phase, the evaluator chose the eight most severe usability errors and problems, according to the evaluator. These eight errors and problems were then sent out to the other three people in order for them to have a look at them. After they had understood the problems all four persons filled out three questions to each problem. The questions were:

- How frequent do you think that the functionality that is affected by the usability problem is used by the user group it applies to? (Frequency of use)
- Out of the times that the functionality is used, how often do you think that this usability problem affects the user in his/her work with the system? (Error frequency)
- How serious do you think that the possible consequences of this usability problem are? (Impact)

These questions were answered on a five-graded scale ranging from zero to four.

When the answers were collected an analysis of them took place. To calculate the severity rating for each problem the following formula was used:

$$AVG(FreqOfUse) * AVG(ErrFreq) * AVG(\text{Impact})$$

Equation 6-1 - Severity rating

Using this formula on each of the eight problems selected for the severity rating, gave the following result:

Table 6.5 - Severity ratings

Problem nr.	Sev. rat.	FOU*EF	I
1	17.88	8.94	2
7	15.47	6.88	2.25
5	13.41	8.94	1.5
6	10.13	4.5	2.25
3	8.44	5.63	1.5
2	3.5	3.5	1
4	3.06	3.06	1
8	0.38	1.5	0.25

These ratings gave the evaluator a basis to work from when performing the cost-benefit analysis. The five problems with the highest severity rating were chosen to undergo a cost-benefit analysis.

6.5 CBA

The cost-benefit analysis is a process where the problems found are being analyzed in a cost-benefit perspective. This means that each problem is going to result in some sort of amount for how much the company can save if the problem is fixed.

The analysis was executed by the evaluator, in collaboration with an AZ staff with good domain knowledge as well as a good economical knowledge within AZ. The analysis was intended to be performed this way, but there was a slight change of plans. The analysis instead did not include the evaluator as much as planned. Instead three other, more qualified, AZ-staff were contacted and they had a teleconference where the problems were discussed and analyzed. The result from the session can be observed below.

Table 6.6 - CBA of StudyAdmin

Problem number	1	7	5	6	3
Reduction of learning time (minutes/new user)	130	130	20	40	65
Speed-up in expert performance (% of time)	2.0%	1.5%	1.5%	1.0%	1.0%
Average time in system per week/user (hours)	9	9	9	9	9
Number of users	1200	1200	1200	1200	1200
Number of new users/year	280	280	280	280	280
Savings that are realizable (%)	33%	33%	33%	33%	33%
Tot. reduction of learning time (hours/year)	200	200	31	62	100
Tot. speed-up expert performance (hours/year)	2851	2138	2138	1426	1426
Cost ($/system hour/user)	80	80	80	80	80
System lifetime (weeks)	156	156	156	156	156
Saved learning costs ($)	16016	16016	2464	4928	8008
Saved performance costs ($)	228096	171072	171072	114048	114048
Total cost reduction ($)	244112	187088	173536	118976	122056

The numbers above is strictly based on assumptions by the three AZ staff at the meeting they held. The first two of the fields, reduction of learning time and speed-up in expert performance, were derived with the severity rating in mind. The time assumptions made in the severity rating procedure formed a basis for the time assumptions made in this cost-benefit analysis.

The reduction of learning time is a measure of how much time a new user would save in training, if the problem were fixed. To calculate these numbers the total training time for each module had to be calculated. The patient matrix is a big focus in training and approximately 26 hours is spent in learning the patient matrix. Problem number one, seven and three are problems within the patient matrix and the timesavings are therefore based on that assumption. The learning times for problem number five and six is four hours.

The speed-up in expert performance is a percentage based best guess in how much time a user, who has finished his training, can save if the problem was fixed. The AZ staff explicitly expressed that this was a very hard number to estimate. But one has to remember that all these numbers are qualified guesses and they are meant to be as accurate as possible. To make all numbers as accurate as possible all previous data collected were used as a basis for the analysis. With all this background information the assumptions made should be very close to the actual numbers.

The average time with the system is taken directly from the mean number of hours the monitors said that they worked with it in the attitude surveys.

The number of users and new users were colleted from a statistical sheet within AZ.

Savings that are realizable is a number Nielsen (1994) has developed. This means that, when the developers fix the problem, the problem does not go away completely. This is because when the developers correct the problem they replace the code with new code that contains problems as well. 33% might seem like a low number, but it means that some parts of the complex problem are corrected at least.

The total reduction of learning time and the total speed-up in expert performance is the corresponding number of hours saved multiplied with the number of users affected. The factor of savings that are realizable is also multiplied to get the correct number of hours.

The total cost reduction provides a number for how much money that could be saved if the problems are fixed in one year. It does not include any numbers for how much the developers will charge for fixing them. The value of each usability problem can be looked as a business market value for each problem. That means the current value of the usability problem.

The system lifetime is a factor that comes in handy, when the actual saving should be calculated. If a system has a short system lifetime, the problems that are to be fixed, have to be cheap to correct and have a high impact.

6.6 Case discussion

The case study was completed as planned with one exception. The CBA was conducted in a slightly different way than planned. Instead of including the evaluator in all of the steps, the three AZ staff took care of the estimates themselves. This should not be a drawback since the evaluator does not have the inside information needed to make the estimations. The downside, to not include the evaluator, can be that the AZ staff could misread the problem report and not analyze the real problem. In this case, an hour-long meeting was held between the evaluator and one of the AZ staff to prevent this.

The expression learning time caused some confusion at first. The definition of what specified a new user was a bit diffuse. When do a user become an expert? In the CBA, the training time was presumed to be 39 hours. First of is three hours of training in a classroom. Then there are monitoring visits where the new monitor has assistance of an expert user. The total time for these accompanied visits are 36 hours.

The heuristic evaluation is a somewhat delicate procedure time wise. It is hard to know when to stop and this was the case in this case study as well. The evaluation stopped when the evaluator thought that he had found enough problems to present. The evaluator also had a time limit for how long he could take on the evaluation. Since one of the main goals for the framework is that it should be performed within 40 man-hours, the heuristic evaluation should not take too long to execute.

6.7 Usability problem report 1 – First summary

This problem report is the first version that the evaluator derived during his heuristic evaluation.

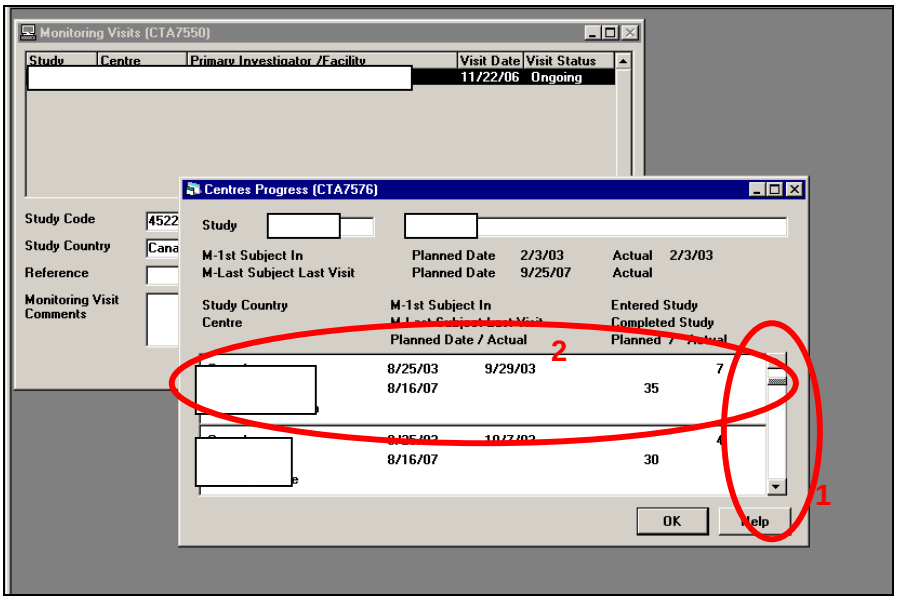

Figure 6.2 - Study/Centres progress

1. Scroll wheel does not work. Long list creates tensions.
2. No dividing lines between the different information fields create confusion.

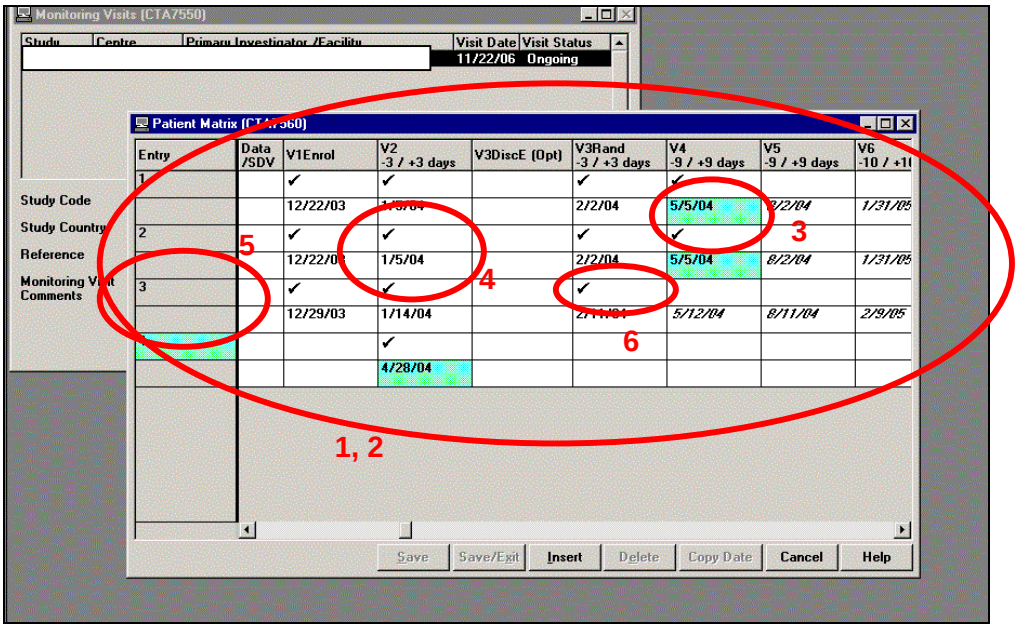

Figure 6.3 - Monitoring/Patient matrix

1. No possibility to change order of the fields. Monitors would like to change in order to see what is most important for them.
2. No dividing lines between different patients create confusion when you scroll to the right.
3. Green colour marks that something has to be checked. The cultural constraint for the colour green is that something seems ok.
4. The check mark above the date creates confusion to what date the mark connects.
5. When you highlight a patient, in order to look at details etc., the highlighting effect is not clear at all. It could be more distinct.

6. It is too easy to accidentally uncheck the box if you double-click. The user is custom to double-click in order to perform his task and double-clicking it should make the ticking of a box.

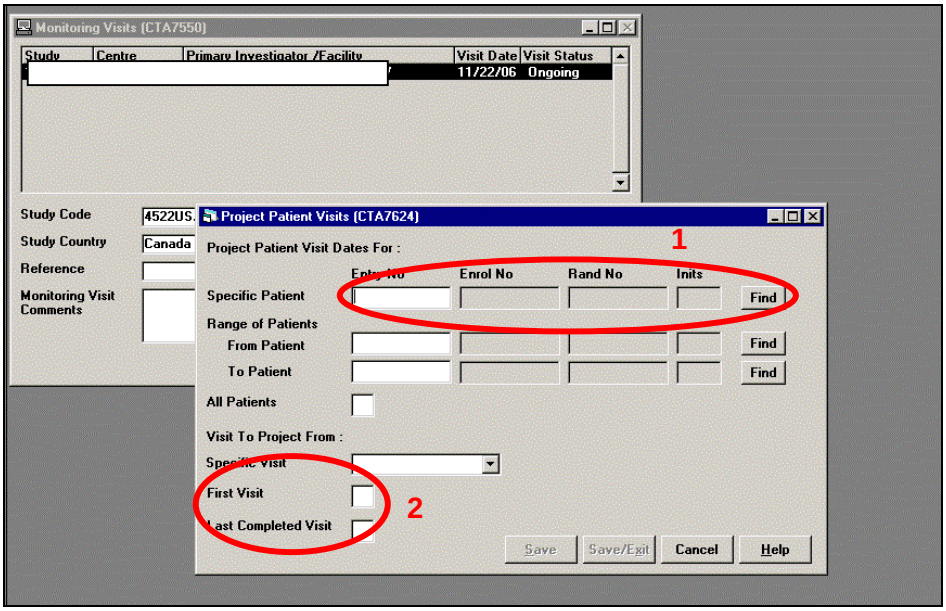

Figure 6.4 - Monitoring/Project Patient Visits

1. When you enter a number in the textbox and press “Find”, you have to enter the number again in the new dialog that pops up. It takes double the effort of what is necessary.
2. Checkboxes should only be used when there is a possibility to enter more than one alternative. In this case only one of the boxes can be ticked at one time and therefore radio boxes should be used instead.

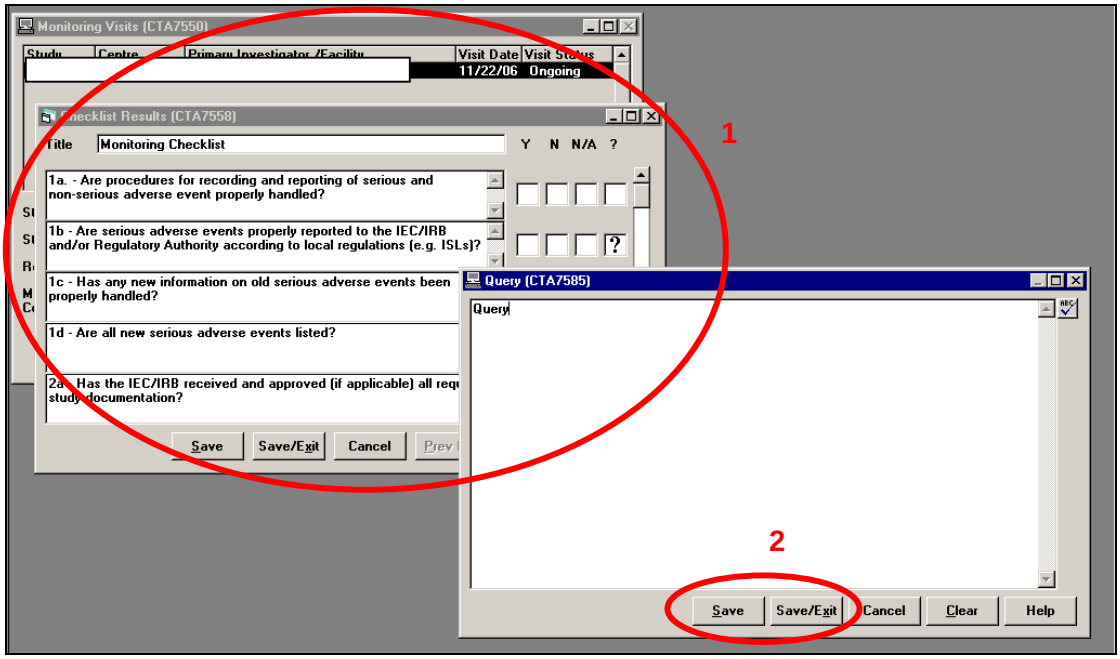

Figure 6.5 - Checklist results

1. Checklist questions sometimes come in the wrong order compared to the crf. This can lead to confusion.

2. If you press “Save” to save your query, the exit button fades and the only option left, in order to exit the dialog, is to press “cancel” or the upper right “x”. This can be comprehended, as the save was not made.

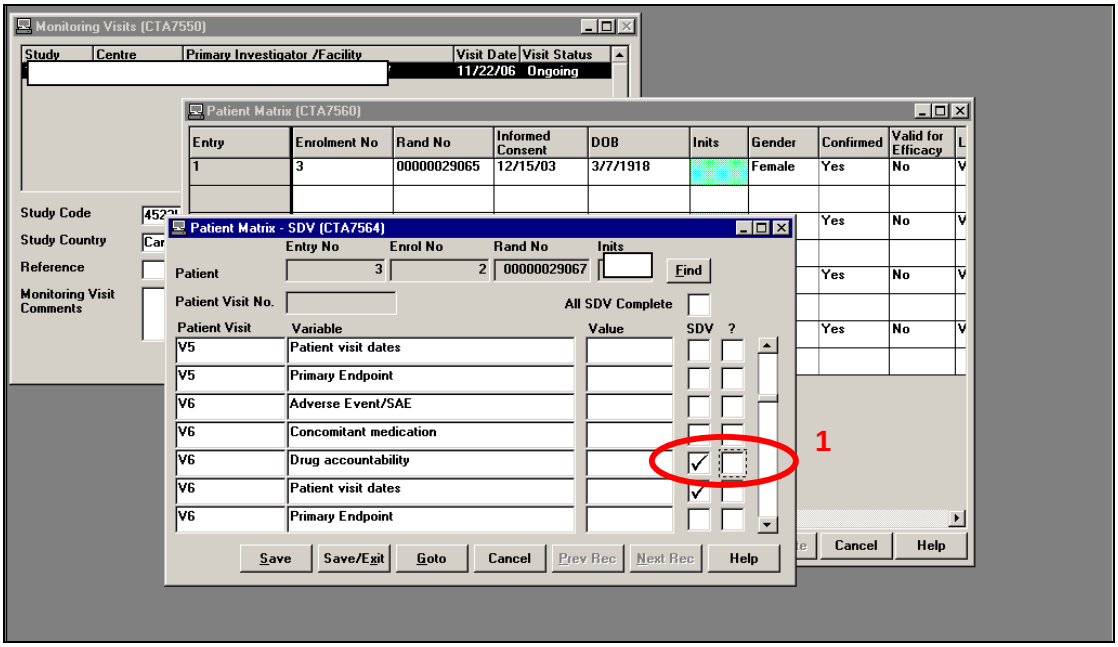

Figure 6.6 - Patient matrix/SDV

1. ? -Mark is not possible to fill out if checkbox is ticked. If it is not possible, it could be made grayish to make a mark that is it not an option at the moment.

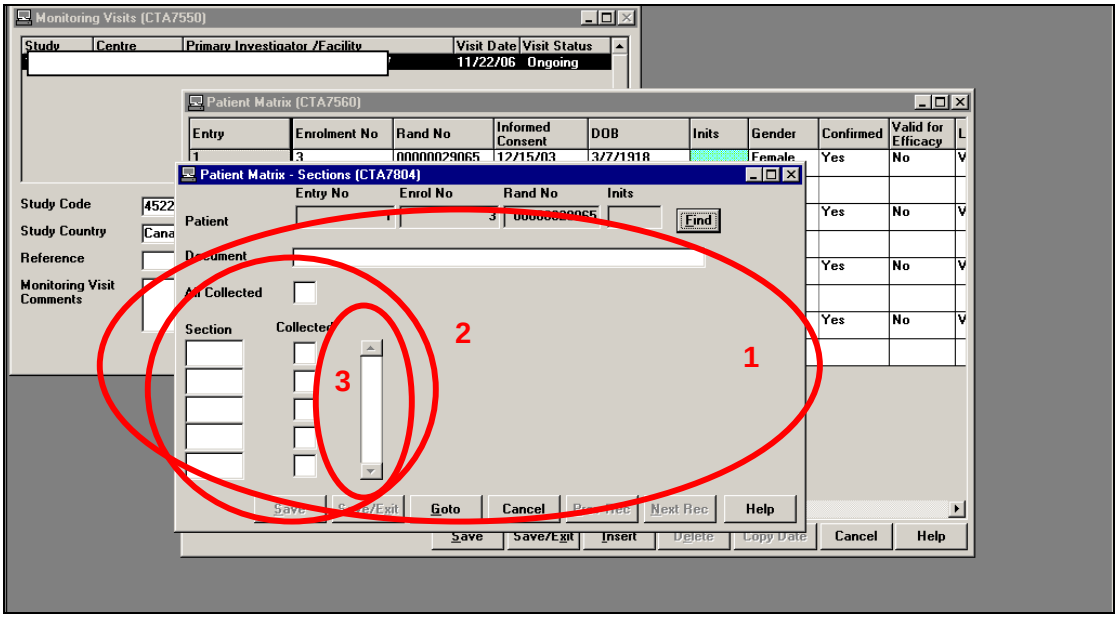

Figure 6.7 - Patient matrix/sections

1. Not possible to edit textboxes, although they appear to be editable.
2. Alignment of textboxes is not very well. It creates unnecessary movement of the eye when you fill them all out. Should be aligned more to the middle if this configuration is to be preserved.
3. If a scrollbar is needed, it should be separated in some way to not confuse the user.

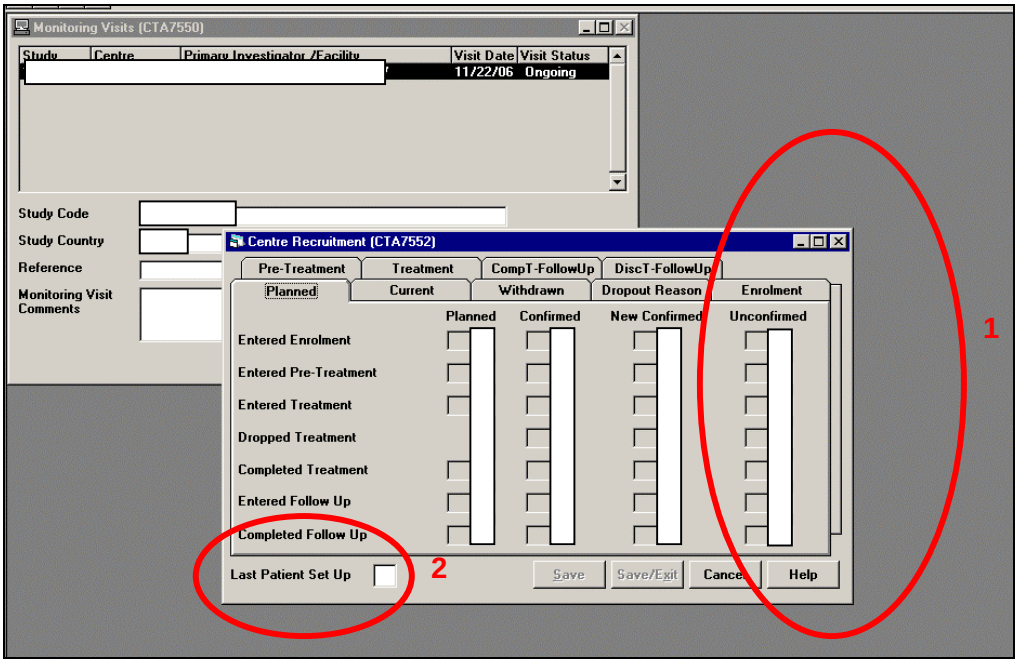

Figure 6.8 - Centre recruitment

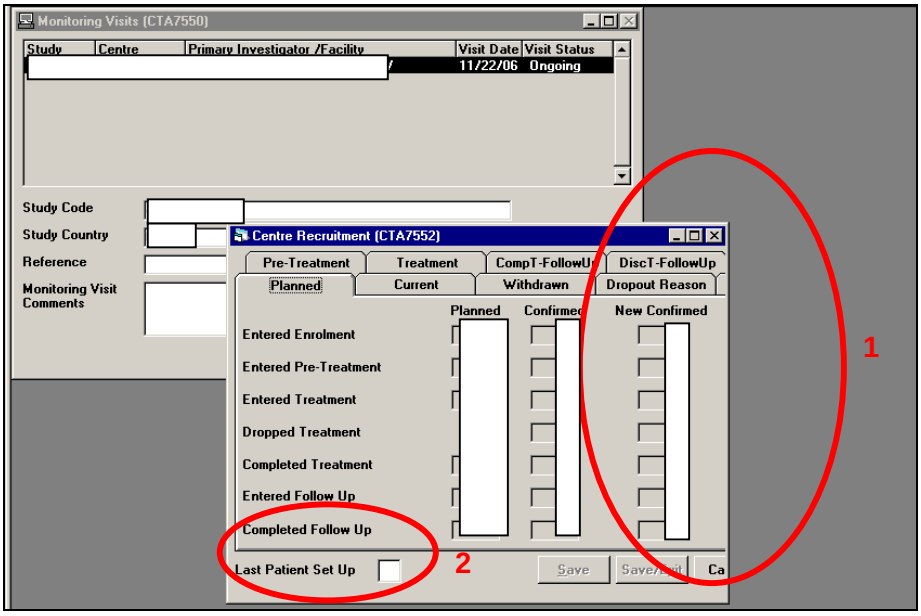

Figure 6.9 - Centre recruitment

1. Resizing the window does not cause scrollbars to appear. Should resizing be an option? This can lead to confusion and information loss.
2. Checkbox is not available and should therefore be grayish.

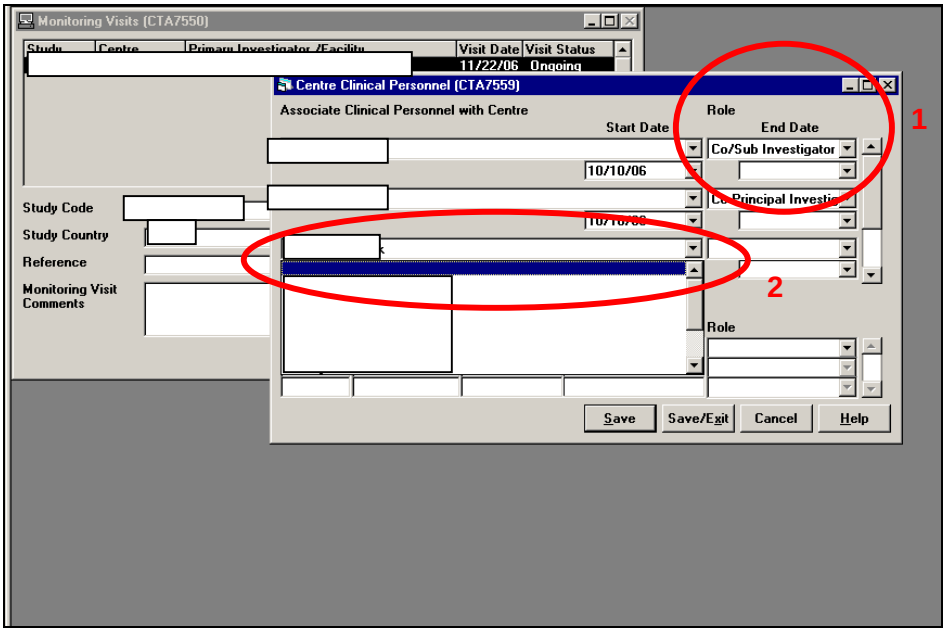

Figure 6.10 - Centre clinical personnel

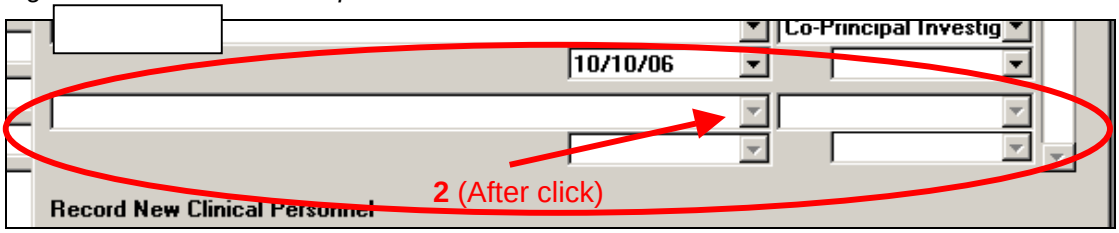

Figure 6.11 - Centre clinical personnel (Detail)

1. The mapping between “Role” and “End Date” and their related boxes is not clear. Could be made clearer if the alignment was different.
2. If the empty box is selected, the whole option of choosing centre clinical personnel is faded and is not longer available. There is no longer a possibility to choose new personnel without saving, exiting and entering the dialog again.

Other known problems of StudyAdmin

A problem for the Swedish monitors can be that Swedes are used to be able to write all dates the same way (YYYY-MM-DD), regardless if it's a date for a meeting or a date of birth. The different standards in StudyAdmin Host and StudyAdmin get confusing and irritating. Especially StudyAdmin Host is considered illogical because of the mix of date formats.

How the patients achieve different states seem to be confusing. A monitor stated that the patient’s first visit has to be monitored before being visible in the inclusion diagram even if an inclusion card has been received from the investigator and the information has been added to StudyAdmin Host. This becomes a problem if the recruitment is in the end phase and a lot of new patients are included. The monitor describes the difference between enrolled patients and patients visible in the inclusion diagram as significant. It can sometimes look like there are 8 patients in a study when the actual enrolled number is 25.

There is much inconsistency between screens in StudyAdmin. A standardized way of designing the screen would help the user recognize and work around problems that may occur.

6.8 Usability problem report 2 – With severity ratings

This problem report is complete with severity ratings. The first five problems were later sent out to be evaluated in a cost-benefit analysis.

Usability problem 1

Severity rating: 18

Frequency of use: 3.25 (3 – Daily, 4 – Several times a day)
Error frequency: 2.75 (2 – Every now and then, 3 – Often)
Impact: 2 (2 – Serious)

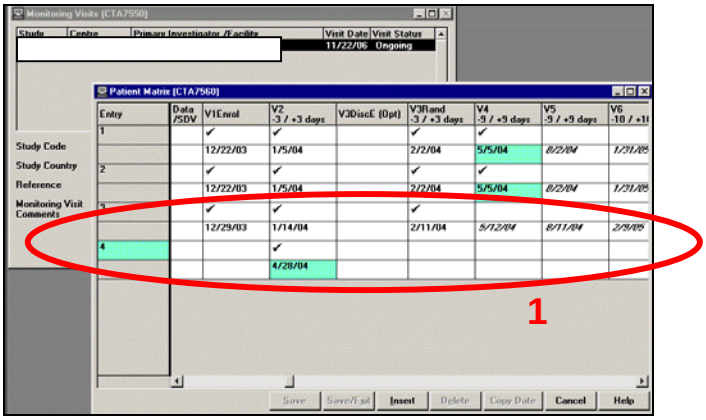

Figure 6.12 - Monitoring/Patient matrix

Dividing lines are missing
This can cause the user to connect wrong dates to wrong check marks and therefore check a field connecting to the wrong patient. If the user does this, the consequence can be that the user thinks that he has checked a visit for one patient, when he has not really. This can lead to further incorrectness in the study progress.

Usability problem 7

Severity rating: 15.5

Frequency of use: 2.75 (2 – Weekly, 3 – Daily)
Error frequency: 2.5 (2 – Every now and then, 3 – Often)
Impact: 2.25 (2 – Serious, 3 – Major)

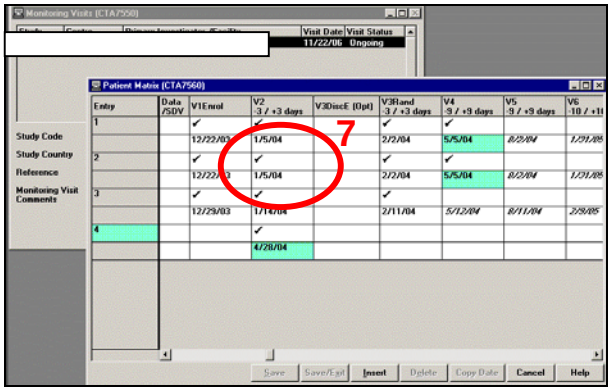

Figure 6.13 - Monitoring/Patient matrix

Check mark above date can create confusion
The check mark is placed above the date in the patient matrix. This can cause the user to think that wrong date is connected to the wrong check mark. If the user does this, wrong patient can be marked as visited.

Usability problem 5

Severity rating: 13.4

Frequency of use: 3.25 (3 – Daily, 4 – Several times a day)
Error frequency: 2.75 (2 – Every now and then, 3 – Often)
Impact: 1.5 (1 – Minor, 2 – Serious)

Different date formats creates confusion
A problem for the Swedish monitors can be that Swedes are used to be able to write all dates the same way (YYYY-MM-DD), regardless if it's a date for a meeting or a date of birth. The different standards in StudyAdmin Host and StudyAdmin get confusing and irritating. Especially StudyAdmin Host is considered illogical because of the mix of date formats.

Usability problem 6

Severity rating: 10.1

Frequency of use: 2 (2 – Weekly)
Error frequency: 2.25 (2 – Every now and then, 3 – Often)
Impact: 2.25 (2 – Serious, 3 – Major)

Information presentation
How the patients achieve different states seem to be confusing. A monitor stated that the patient's first visit has to be monitored before being visible in the inclusion diagram even if an inclusion card has been received from the investigator and the information has been added to StudyAdmin Host. This becomes a problem if the recruitment is in the end phase and a lot of new patients are included. The monitor

describes the difference between enrolled patients and patients visible in the inclusion diagram as significant. It can sometimes look like there are 8 patients in a study when the actual enrolled number is 25.

Usability problem 3

Severity rating: 8.4

Frequency of use: 2.5 (2 – Weekly, 3 – Daily)
Error frequency: 2.25 (2 – Every now and then, 3 – Often)
Impact: 1.5 (1 – Minor, 2 – Serious)

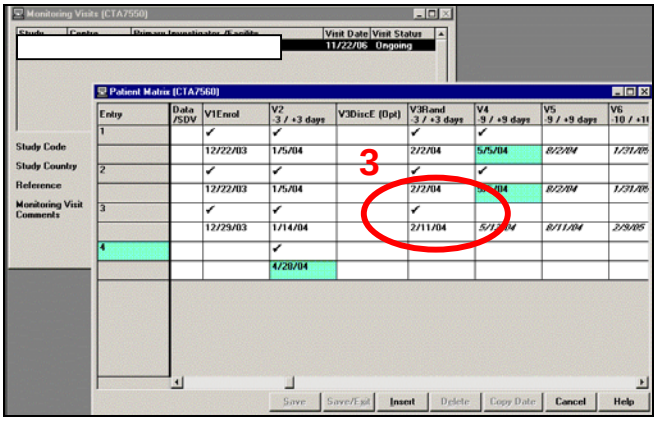

Figure 6.14 - Monitoring/Patient matrix

Double-clicking causes the box to stay unchecked
This phenomenon can trick the user to think that, since he double-clicked the check mark, it should stay checked. This behaviour is against the regular behaviour of Windows® where double-clicking indicates opening or marking an object.

Usability problem 2

Severity rating: 3.5

Frequency of use: 1.75 (2 – Weekly, 3 – Daily)
Error frequency: 2 (2 – Every now and then)
Impact: 1 (1 – Minor)

Inconsistency
There is much inconsistency between screens in StudyAdmin. A standardized way of designing the screen would help the user recognize and work around problems that may occur.

Usability problem 4

Severity rating: 3

Frequency of use: 1.75 (1 – Monthly, 2 – Weekly)
Error frequency: 1.75 (1 – At rare occasions, 2 – Every now and then)
Impact: 1 (1 – Minor)

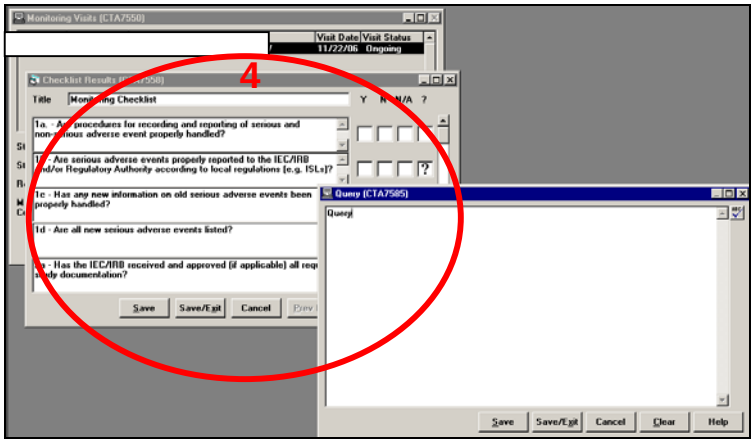

Figure 6.15 - Checklist results

Checklist questions can come in wrong order to what they are in the CRF.
As pointed out by one of the monitors, the checklist questions sometimes come in the wrong order to what they are presented in the CRF. This can create a problem for the user and cause irritation and confusion.

Usability problem 8

Severity rating: 3

Frequency of use: 1.5 (1 – Monthly, 2 – Weekly)
Error frequency: 1 (1 – At rare occasions)
Impact: 0.25 (0 – Not an issue, 1 – Minor)

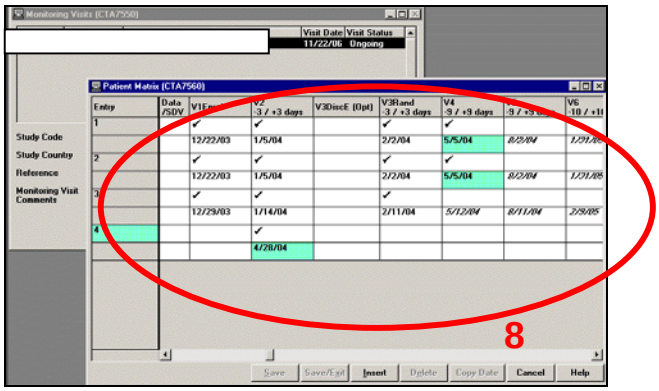

Figure 6.16 - Monitoring/Patient matrix

Reorder as you like is not possible
In the patient matrix it is not possible to create any order of the information fields. If the user can rearrange in what order the fields should be presented, higher user satisfaction can be achieved. Different users have different preferences to what should be seen at first sight. I.e. different users think that different things are important for their own sake.

6.9 Usability problem report 3 – With CBA

This problem report is the one that presents all problems, with corresponding severity ratings, for traceability, and corresponding cost-benefit analysis. The numbers presented for each problem was carefully derived and can serve as a basis for further error management.

Usability problem 1

Reduction of learning time (minutes/new user)	130
Speed-up in expert performance (% of time)	2,0%
Average time in system per week/user (hours)	9
Number of users	1200
Number of new users/year	280
Savings that are realizable (%)	33%
Tot. reduction of learning time (hours/year)	200
Tot. speed-up expert performance (hours/year)	2851
Cost ($/system hour/user)	80
System lifetime (weeks)	156
Saved learning costs ($)	16016
Saved performance costs ($)	228096
Total cost reduction ($)	244112

Severity rating: 18

Frequency of use: 3.25 (3 – Daily, 4 – Several times a day)
Error frequency: 2.75 (2 – Every now and then, 3 – Often)
Impact: 2 (2 – Serious)

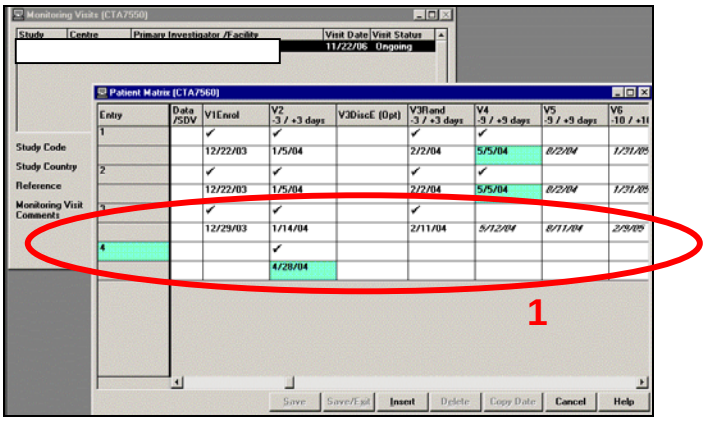

Figure 6.17 - Monitoring/Patient matrix

Dividing lines are missing

This can cause the user to connect wrong dates to wrong check marks and therefore check a field connecting to the wrong patient. If the user does this, the consequence can be that the user thinks that he has checked a visit for one patient, when he has not really. This can lead to further incorrectness in the study progress.

Usability problem 7

Reduction of learning time (minutes/new user)	130
Speed-up in expert performance (% of time)	1.5%
Average time in system per week/user (hours)	9
Number of users	1200
Number of new users/year	280
Savings that are realizable (%)	33%
Tot. reduction of learning time (hours/year)	200
Tot. speed-up expert performance (hours/year)	2138.4
Cost ($/system hour/user)	80
System lifetime (weeks)	156
Saved learning costs ($)	16016
Saved performance costs ($)	171072
Total cost reduction ($)	187088

Severity rating: 15.5

Frequency of use: 2.75 (2 – Weekly, 3 – Daily)
Error frequency: 2.5 (2 – Every now and then, 3 – Often)
Impact: 2.25 (2 – Serious, 3 – Major)

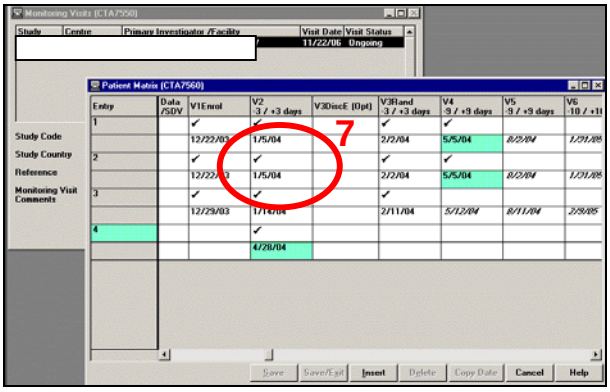

Figure 6.18 - Monitoring/Patient matrix

Check mark above date can create confusion
The check mark is placed above the date in the patient matrix. This can cause the user to think that wrong date is connected to the wrong check mark. If the user does this, wrong patient can be marked as visited.

Usability problem 5

Reduction of learning time (minutes/new user)	20
Speed-up in expert performance (% of time)	1.5%
Average time in system per week/user (hours)	9
Number of users	1200
Number of new users/year	280
Savings that are realizable (%)	33%
Tot. reduction of learning time (hours/year)	31
Tot. speed-up expert performance (hours/year)	2138.4
Cost ($/system hour/user)	80
System lifetime (weeks)	156
Saved learning costs ($)	2464
Saved performance costs ($)	171072
Total cost reduction ($)	173536

Severity rating: 13.4

Frequency of use: 3.25 (3 – Daily, 4 – Several times a day)
Error frequency: 2.75 (2 – Every now and then, 3 – Often)
Impact: 1.5 (1 – Minor, 2 – Serious)

Different date formats creates confusion
A problem for the Swedish monitors can be that Swedes are used to be able to write all dates the same way (YYYY-MM-DD), regardless if it's a date for a meeting or a date of birth. The different standards in StudyAdmin Host and StudyAdmin get confusing and irritating. Especially StudyAdmin Host is considered illogical because of the mix of date formats.

Usability problem 6

Reduction of learning time (minutes/new user)	40
Speed-up in expert performance (% of time)	1.0%
Average time in system per week/user (hours)	9
Number of users	1200
Number of new users/year	280
Savings that are realizable (%)	33%
Tot. reduction of learning time (hours/year)	62
Tot. speed-up expert performance (hours/year)	1425.6
Cost ($/system hour/user)	80
System lifetime (weeks)	156
Saved learning costs ($)	4928
Saved performance costs ($)	114048
Total cost reduction ($)	118976

Severity rating: 10.1

Frequency of use: 2 (2 – Weekly)
Error frequency: 2.25 (2 – Every now and then, 3 – Often)
Impact: 2.25 (2 – Serious, 3 – Major)

Information presentation
How the patients achieve different states seem to be confusing. A monitor stated that the patient's first visit has to be monitored before being visible in the inclusion diagram even if an inclusion card has been received from the investigator and the information has been added to StudyAdmin Host. This becomes a problem if the recruitment is in the end phase and a lot of new patients are included. The monitor describes the difference between enrolled patients and patients visible in the inclusion diagram as significant. It can sometimes look like there are 8 patients in a study when the actual enrolled number is 25.

Usability problem 3

Reduction of learning time (minutes/new user)	65
Speed-up in expert performance (% of time)	1.0%
Average time in system per week/user (hours)	9
Number of users	1200
Number of new users/year	280
Savings that are realizable (%)	33%
Tot. reduction of learning time (hours/year)	100
Tot. speed-up expert performance (hours/year)	1425.6
Cost ($/system hour/user)	80
System lifetime (weeks)	156
Saved learning costs ($)	8008
Saved performance costs ($)	114048
Total cost reduction ($)	122056

Severity rating: 8.4

Frequency of use: 2.5 (2 – Weekly, 3 – Daily)
Error frequency: 2.25 (2 – Every now and then, 3 – Often)
Impact: 1.5 (1 – Minor, 2 – Serious)

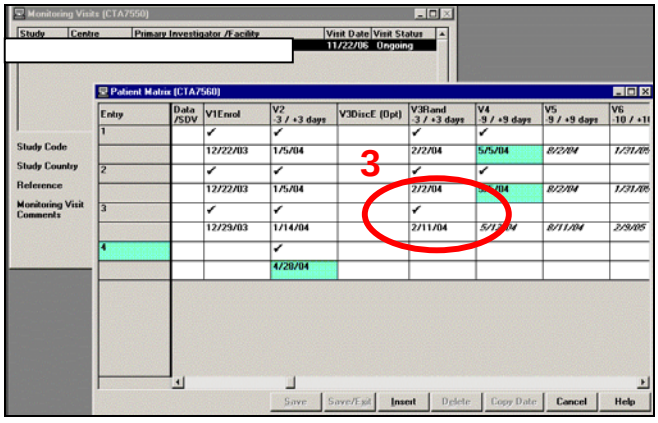

Figure 6.19 - Monitoring/Patient matrix

Double-clicking causes the box to stay unchecked
This phenomenon can trick the user to think that, since he double-clicked the check mark, it should stay checked. This behaviour is against the regular behaviour of Windows® where double-clicking indicates opening or marking an object.

---

7 Framework discussion

This chapter contains a discussion on how the final framework should be put together. The discussion is based on the information in the theory chapter, the first framework proposal and how the first framework proposal worked in practice. Based on the case studies, the pros and cons of the framework will be discussed, as well as the ideas and opinions that the evaluators and other staff members expressed during the case studies and in discussions concerning the framework. The conclusions in that are drawn in this chapter will form the basis for the final framework that will be presented in the following chapter.

7.1 Background studies

The background studies are supposed to be the first part of the framework. This means that the background studies are to result in that the evaluator collects all needed information to continue his process. There are four main activities in the background studies; Start up meeting, documentation studies, interviews and field study / observation. These four activities have been chosen through the case studies and have been proven to fit the purpose well.

The start up meeting is held to give the evaluator the first much needed information about the system and its users. The case studies shows that the start up meeting can also be used to find out a bit more about the organization and which users to contact to get further information.

The documentation studies are included because it is necessary for the evaluator to acquire an understanding of the system and its functionality in order to perform a heuristic evaluation. The documentation studies can be performed in different ways. In the first case study the documentation study was conducted through an e-learning session. This proved very handy and can be recommended in this study. In the second case study the documentation study was achieved through reading a user's manual. This alternative can produce as good results as the e learning, but it has its drawbacks. One is that it is hard to read and then understand the information directly. Learning by doing is one of the most effective ways of learning. An additional problem with reading documentation is that it is much more time-consuming to gain the same knowledge as you would through the e-learning.

It is important for the evaluator to gather as much input as possible from system users, however the evaluator has to make a judgment of how many interviews that are reasonable to spend time on. The framework is under a certain time pressure and every time consuming activity has to be motivated. The interviews can help very much in giving the evaluator a better understanding of the system, but every new interview does not have to contribute as much. Different users should be interviewed, but they have to be carefully picked out in order to get a broad field of user types.

The field study / observation have to be performed in order for the evaluator to see how the system is really used. In both case studies, the observation was conducted at a care center where a monitor did his usual work and the evaluator watched. It can be argued that the monitor did not do his job as usual, because when working with a bystander, the workflow might be affected slightly. The monitor may pay extra attention not to make mistakes in front of another person and can therefore affect the observation a bit. The main purpose of the visit was though to see how the system is used in its proper context. This aspect was not affected by the fact that the evaluator may have disturbed the monitor in his ordinary work.

Validity
The background studies are an important part of the framework, since the groundwork is laid in this phase. The knowledge gained in this phase are further refined during later stages of the framework. It is also important to understand the systems users. Problems with the usability of a system only appear when a user is interacting with it. With a good user analysis, the evaluator has a much easier time deriving the use cases and analyzing the system better.

The evaluator can affect the mix of different methods of acquiring the information needed to continue into the next phase of the framework. It is important though that the results are analyzed and evaluated in terms of validity. If the information is not valid, the rest of the procedure can be misleading. To base a decision on one user's opinion is dangerous, because that opinion may not represent the rest of the system user's. This is a factor that relies on the evaluator, and on his experience, and the judgment he makes, whether to continue the background studies or not, has to be made on fair grounds. An insight in the organization is needed in order to make the correct decision.

The background studies give the evaluator a first glance at the system and its possible usability problems. The problems found in this phase are not complete, but should be documented for further evaluation in later stages of the framework. The problems give the evaluator a chance to see what the users find difficult and can lead to important discoveries in a flow analysis. The case studies show that many of the problems that were rated as the most serious were found during the background studies phase. Some of the problems found here were rated as not important at all as well. This shows that it is important not to make a judgment based on one user's opinion. These problems were often found, not as real defined problems during the observation rather than that they were talked about during the small talk during the observation and the interviews.

Reliability
The issue regarding what users to contact and how many to talk to, is an important one. As shown above, the number of observations or interviews that are conducted can easily affect the quality of the problems derived. A few users mean that the problems may not represent what the majority of the users think. A problem that is important to one person may not be important to the rest of the population as shown in usability problem 8 in *Case study – StudyAdmin: Usability problem report 2 – With severity ratings.* This shows a problem found in the monitoring visit where the monitor found this problem severe, but when the problem was analyzed in the severity rating process, it clearly showed that it was an isolated event. To base the use cases on a small selection of interviewee's answers is not a good way of deriving use cases. The evaluator has to have enough experience to sense this and correct it by including more people in the background studies.

The problem with the reliability at the monitoring visit concerning that the monitor might not show his real working procedure because of the disturbance of the evaluator is not a very important aspect. It is definitely worth considering at, but should not be taken into great considerations. The presentation is still made by the monitor and the evaluator gets a glimpse of what problems might occur. The workflow still has to be followed by the monitor and can easily be studied by the evaluator. Another aspect to study during the observation is the working conditions the monitors have to face. Ergonomic factors can affect the user personally and problems, such as poor light, can certainly affect the user in a negative way. In both

case studies this was not considered a problem and the possible side effects were taken out of considerations.

Conclusions
As our case studies show, the background studies made were sufficient for their purpose. The process proposed in the first framework proposal, start up meeting, documentation studies, interviews and field study / observation, proved to be good. The time spent on the different stages in the background studies has to be decided by the evaluator and be based on his/her experience a decision has to be made whether it is enough or if it have to continue a while longer.

The goal of the evaluation has to be decided in the beginning of the project. A first calculation of what the profit for the company can be should be made.

The right user types have to be selected for interviews and observations. A project with a large evaluation budget might want to spend some more time on the interviews and the observations. The more people that are interviewed, the better view of the system the evaluator gets. It is also important to combine the interviews with documentation studies. The case studies show that the e-learning method is the best way to gain an understanding of the system in shortest possible time.

The interviews have to be well prepared. Questions can be written in advance and the purpose of a specific interview has to be defined. A semi-structured way of interviewing is the best way of getting the most information possible. The interviews can, with advantage, be held through a telephone conference. This means that you can interview people from other countries and the time does not have to be a major factor. A telephone conference can be held at anytime and does not have to be very formal. The main goal is to get the correct information. Possible interviewees can be; staff with a support function, system administrators, regular users, staff with system responsibilities and system owners.

The scenarios chosen for the evaluation is based on the information collected in the background studies. This means that the information has to be of a certain quality and has to be valid. It is important to specify how to choose the critical use cases. The most used functions; the functions that are most important not to be erroneous and the most error-prone functions have to be included in the later analysis.

7.2 Evaluation

The evaluation is based on the findings of the background study and the experience of the evaluator. It is important for the evaluator to use the critical workflows and the possible usability problems found in his/her evaluation. The evaluation method suggested, is two parted. The first part, the attitude survey, is to be considered as a more iterative process that is supposed to be a part of the ongoing usability work at AstraZeneca. It can be used to get the attitude towards the usability of a system. The second part, the usability evaluation, is very much up to the evaluator to sort out. How it is performed is up to the evaluator, but the result has to be a specified list of usability problems ready to be prioritized in the next phase.

7.2.1 Attitude Survey

The attitude survey is based on a known scale called system usability scale (SUS) (Brooke, J. 1986). It provides the attitude towards the usability of a system on a scale that ranges from zero to 100.

Validity
The SUS survey is used in this framework because it is hard to get valid responses regarding the users' opinions when the evaluator himself/herself asks them. It is also important because the attitude is such an important part of the term usability. A bad attitude towards the usability of a system can often be interpreted as that there are problems with the system.

In this framework the SUS is going to be used as a pointer to show that changes made have an impact on the users. It is important to show that there are results in the attitude when the company invests money to improve the usability. Since the scale is already made, there does not have to be much time spent on this activity by the evaluator. The only thing the evaluator has to do is to send out the surveys and when the surveys are returned the evaluator has to calculate the usability score.

Reliability
The case studies show that the usability scale works for the purpose. In the background studies it was expressed that the interviewees thought that the usability in StudyAdmin was worse than in StudyCapture. This was later further illustrated by the results from the attitude survey, where StudyAdmin got a lower score.

The scale cannot be directly translated into the overall usability of a program. Just because a program gets a low score in the SUS, it does not mean that the overall usability is bad. If the score is low it can imply that there are problems with the program, but you cannot tell how much problems or how severe they are.

The scale is best suited for comparison between different versions of a program. In this case you know that the only different thing between what is tested is the problems that are fixed. If the attitude has improved, you can interpret that, as the changes made were successful. The range that the SUS use ranges from zero to 100. 50 can be looked as a neutral result. This means that any result over 50 is a positive attitude towards the usability. And any result below is considered a negative attitude. It is important to select the proper survey participants. The primary users should be selected because it is their opinions that matter the most.

The number of participants for the survey has to be decided by the evaluator at the start of the evaluation phase. 15 should be enough to make general conclusions on what the population thinks of the product. If more than one user group are to be evaluated; 15 survey responses per user group is needed for enough reliability.

Conclusions
The attitude surveys should be a part of the framework since the attitude towards usability is an important aspect of the term usability. It should not take too much time from the time frame set for the evaluation. Much of the work with the attitude survey is already made in advance.

The evaluator has to be careful in his/her interpretation of the results from the survey. You cannot draw to many conclusions on the results, but it can help you make estimation on how many usability problems there are and how severe they might be. How positive / negative the results are, are almost impossible to tell at first. The comparison of different results requires experience and previous results to compare with. At first you can only tell if there is a positive or negative attitude.

The survey can and should be sent out as soon as the user groups are defined in the background studies. The sooner you send them out, the sooner you will get them back for analysis.

The main use for the attitude survey is to measure the difference in attitude between different versions of a program. The results can help show that it is worth spending money on usability.

7.2.2 Usability evaluation

The usability evaluation is the phase where the evaluator produces a usability problem report with specified usability problems. The framework is designed so that the evaluator can perform it the way he/she likes as long as a problem report is the result. It is important to base the evaluation on the findings from the background studies.

Validity

The evaluation phase is where the evaluator finds and specifies all problems. The evaluation is based on the findings from the background studies and the critical flows and the usability problems found there are further investigated. It is important that the evaluator follows the specified time frame since a usability evaluation does not really have a natural end point.

It is important not to spend too much time on the problem report. The problem report is supposed to function as help for the evaluator and therefore does not have to be perfectly made. The time frame for the framework is tight and the report writing process is not worth spending too much time on. The usability report is meant for the evaluator himself and future evaluators and is supposed to aid them with the traceability for the usability problems.

The evaluation itself can be split into two parts. The first takes care of the critical flows. These are important to evaluate because they have most impact on the usability of the program. The second part can be a more general evaluation, where the evaluator wanders around in the system trying to find random problems by analyzing each screen he conquers.

Reliability

It is important to know that different evaluators will find different problems. How good the evaluation is going to be, often depends on the evaluator's experience. The experience an evaluator needs in order to perform a usability evaluation is up to the evaluator himself/herself. If the evaluator thinks he/she can do it and the project lead thinks so too; go for it. The results that come out from a usability evaluation made by a poor evaluator can still be used for further evaluation. The difference is that the better evaluator might find a higher number of problems and more critical problems. A more experienced evaluator might also be able to help more with the further analysis of the problems.

Research shows that more than one evaluator is preferable for good reliability. But the given resources from AstraZeneca, the framework will only use one. If it should be possible to extend the budget for the evaluation it is highly recommended to increase the number of evaluators.

The short amount of time given for the evaluation could get fewer usability problems, but it is the time frame that sets the boundaries. If more time were given, more usability problems would be found.

To save time it is important to base the evaluation on the findings from the background studies. If the focus is correct from the start, no extra time is spent on finding the most critical problems. This is because the interviews with the real users, from the background studies, give the evaluator an early focus.

Conclusions
The evaluator should do both a study of the critical flows and a study of the more general usability problems. The case studies show that the most severe usability problems come from a combination of the two parts.

It is important to include the problems found in the background studies in the evaluation process. Those problems come from real users and may not be found by the evaluator.

A problem report should be written in such a way that not too much time is spent on it, but it should still be sufficient enough to be used in a further evaluation. It should be used as a reference for traceability and track keeper for further evaluation.

The limitations of the evaluation have to be based on the budget. The number of evaluators and time given should be decided at the start of the project.

The evaluation process is a very flexible and adaptable process, where the evaluator can do as he/she chooses. The only real demand there is on the evaluation process is that it is supposed to result in a usability problem report. If more time is given, a more thorough evaluation can be made.

7.3 Usability severity rating

The usability severity-rating phase of the framework is there to prioritize and select the most important problems that have been found in the usability evaluation. The input to this phase is a usability problem report from the evaluation phase. In the case studies the severity rating process was conducted through letting a number of different types of system users rate each usability problem according to frequency and severity. The results were recorded through a survey and summarized.

Validity
Since the time that is spent on usability evaluation is limited, it is of high importance that there is a way to prioritize which usability problems that should be further evaluated and investigated. The usability severity ratings aid the evaluator in the prioritization process as well as provide a basis for the CBA discussion that follows. The severity rating process also provides the evaluator with input from other staff members, which reduces the risk that the prioritization becomes one-sided and that the wrong usability problems get chosen for further analysis. Documenting the summarized severity ratings for each problem also helps motivate the selection process and enables other staff member to trace how the prioritization was made.

The most time-consuming activity in this phase is specifying the usability problems so that they can be understood and rated by others. This specification would be necessary even if the rating process were not present, since the problems need to be specified at some point if they are to be corrected. If a web-survey is prepared with

the questions that are used to rate the problems the process should not take very much time. Summarizing the severity ratings and documenting them is also relatively easy since it can be done according to a predefined formula.

Reliability

In order for the usability ratings to be useful, the ratings and the prioritization that follows must be reliable. However, good results can come out of the evaluation even if the prioritization is not 100 percent accurate. Since the continued analysis of the most serious usability problems result in an estimate of the costs associated with each one of them, the estimate is not affected by which problems that are selected. There is always a possibility of reviewing the selection process and selecting more problems.

According to Nielsen (1994), severity ratings that are set by one evaluator are not trustworthy. Nielsen argues that the reliability of the ratings increase with the number of evaluators that make severity estimates. Since one of the requirements on the framework is that it should be possible to conduct an evaluation in around 40 man-hours, there are not enough resources to bring in more than one evaluator. During the case studies, key users and staff members with insight to the systems were asked to make the severity estimates in substitute for the extra evaluators. This strategy is not documented in any prior research that has been described in the theory chapter, however based on the case study results the strategy was a success. Since the timeframe for an evaluation using the framework is limited the evaluator does not have the time to gain extensive expertise on details of the system. Therefore severity ratings that are made by personnel that have more details on how the system is actually used are more likely to be accurate judgments on how frequently the problem consequences appear and the impact that the consequences have on the workload. However it is important that different user groups are represented in the severity rating process so that the problems that are most important for one user group do not get over- or underrepresented in the further evaluation.

Another issue for the reliability of the severity ratings is the selection and specification process that precedes the rating. The prioritization where the evaluator selects and specifies some of the usability problems that were found in the evaluation phase, is the one part of the framework process were a selection occurs that can not be traced in a document. This selection has to be made solely based on the evaluator's knowledge of the system and usability. The reliability of this selection can be improved through a more experienced evaluator or through selecting more usability problems to be rate in the severity rating process. Furthermore if the rating is going to be trustworthy, the evaluator has to make an effort to describe the usability problems consistently in the documentation so that one error is not perceived as worse than another based on the specification text. Another issue that came up during the case studies was how specific or general the description of the usability problems should be. Two approaches were tested in the two case studies. In one approach the problems were described specifically and in detail, in the other they were summarized into larger groups and described generally. The conclusion was that no approach was better than the other, the evaluator needs to consider this and make a decision on how general the description can be if it is to be understood by other staff members and developer, and how specific it can be if it is going to be important and not a small part of a problem.

Another issue of high importance is the selection of people that are to partake in the rating process. The people that partake should not be the same people that pointed out the problems in the background study since they might be biased towards the

problems that they pointed out. In prior studies (Nielsen, 1994) the relation between the number of evaluators and the reliability of the ratings has been investigated. This research indicates that between five and ten evaluators gives the best efficiency and reliability. Five evaluators should be the minimum since it gives a reasonably high reliability, which increases slowly when the number approaches ten. To gain an overview of the opinions of different user groups the staff that is selected should come from a variety of groups.

The interpretation of the results from the rating activity is just as important as the preparations. If the results are to be reliable, the individual ratings cannot be used on their own. The ratings for each question should be averaged for all participants. The results should then be multiplied to reach the final severity rating for the problem. This approach reduces the risk that wrong choices are made as a result of the method of calculating the severity rating. The averages for each question can then be used in the CBA phase and the rating can primarily be used for prioritizing the problems. In connection to case studies a diagram with all possible results from the rating of a problem was analyzed and found to be reasonable. However as the severity rating is used for more systems the weighing of the three factors could be adjusted if a need arises.

After the severity rating is calculated, the problems with the highest rating should be selected for further analysis in the CBA phase. There is no natural limit on the lowest rating that should be further analyzed; instead this decision should be made each evaluation. In the case studies it was found that the line could easily be found since there, at some point, was a large difference between two numbers. In the StudyCapture case, five problems out of eleven were selected; the lowest score that was selected was 9.375 and the score that came after that was a four-way tie at 5. Therefore the natural line was drawn at 9.375.

Conclusions

Based on the validity and reliability arguments above, the following paragraphs give a few pointers to important aspects of the severity-rating phase.

The evaluator should perform a first prioritization and specification of the problems that he/she thinks are valid to bring into the severity rating process. Through this selection a usability expert makes the first prioritization.

The staff members that estimate the severity ratings should be selected from primary user groups and other system stakeholders such as super users, system owners and administrators. In order for the results to be reliable there should be at least 5 people rating the problems. The evaluator and staff members that have been a part of the background study and reported problems should not be part of this activity. The results should then be specified in a document to support the CBA and enable tracing of the final results.

The extent of the severity rating can be adjusted based on the scale of the evaluation and the budget. The reliability of the evaluator's first prioritization can be improved by selecting more problems for the rating activity. Improvements in the reliability of the severity scores can be made through letting more staff members set severity ratings.

7.4 Cost-benefit analysis

The cost-benefit analysis (CBA) has the purpose of calculating the costs and the benefit market associated with each usability problem. The usability problems that have been found to have the highest severity rating should be chosen for analysis.

Validity

The CBA is needed to support decisions concerning which usability problems that should be fixed as well as to motivate a usability effort towards the rest of the organization and the system developers. If the usability evaluation is not motivated through a CBA and the costs and the benefit market is calculated there is a much larger possibility that the problems remain unfixed and that the effort has been wasted. The CBA can be used as information in Business cases in the organization and therefore connect the usability effort to other calculations and decisions concerning the system.

If the results from the usability evaluation are questioned, it is also important to have a standardized and traceable way of conducting the CBA. In the case studies two different approaches to the CBA were tested. In one, an estimate of what could be saved through fixing the usability problems was made, in the other, an estimate of he benefit market for each problem was made. Discussions resulted in the favoring of the approach were the benefit market was calculated. The argument was that estimates of the savings, without having a proposal of the solution to a problem, would be extremely inaccurate and therefore should not be made. Calculations of the benefit market for a problem on the other hand, estimate the possible gains if the problem was fixed completely. Those figures can be used to make decisions on further investigations and are also much more likely to be correct than the estimated savings mentioned above.

Even though the estimates may not be completely reliable, the effort made is well motivated since it is not important that the estimates are 100 percent accurate. If the estimates can give an indication to the savings that can be made, they can be used as background for decision-making concerning the system.

Reliability

For the cost-benefit analysis to be reliable it is important to have good background information and input to bring to the meeting. Based on the case studies it was concluded that the severity ratings provided a strong basis for the estimates during the CBA meeting. However it is important that the evaluator stresses that the severity ratings have to be considered during the meeting and that it is not only the participants view on the problems that counts when making the estimate. The severity ratings provide the CBA meeting participants with the opinions of others with regards to the usability problems and can therefore help improve the reliability of the CBA.

It is also important to choose the right participants for the meeting if the results are to be reliable. The evaluator should lead the meeting and provide usability expertise, in addition to that two or more experts on the system with knowledge of the organization and system details and tasks should participate. This combination of people should be able to judge each problem according to frequency and the time spent if it occurs.

Based on the case studies it was also concluded that it would benefit the reliability of the results if each problem was analyzed for its possible consequences and the estimates were based on those consequences. The StudyCapture and the

StudyAdmin case study had two different approaches to this and it was concluded that if an analysis of the consequences of each problem was made, it was easier for the CBA meeting participants to make their estimates, more reliable results were achieved and the process could be traced better. The consequence analysis could with advantage be performed using sketches of the possible consequences.

During the case studies it was also concluded that the estimates of problem consequences in expert performance and possible reduction of learning time would serve the purpose well. As mentioned in the validity paragraph it was concluded that the best approach for the CBA was to estimate the benefit market for each problem and not make estimates of what actually would be saved if the problem were fixed. The reliability of the estimates will also be improved if the calculations stops at the benefit market, since the estimate of how well a usability problem can be fixed is eliminated. The reliability of the estimates that are made in at the CBA meeting can further be improved by defining upper and lower bounds for the benefit market.

An issue with the reliability of the CBA estimates is that the estimates that were made during the case studies were not tested and confirmed to be correct. To ensure that the method of the CBA is reliable the estimates should be put to the test through actually improving the functionality that is affected by the usability problems.

Conclusions
Based on the arguments above the following conclusions were made about the CBA.
The evaluator is responsible for providing the required background material for the CBA meeting. The results from the meeting are highly dependent on the quality of the background information. Furthermore the staff members that are selected to participate need to possess knowledge of both system details and the organization surrounding the system.

The CBA should be performed through an analysis of the possible problem consequences. The goal for the evaluation should be to make an estimate of the benefit market for fixing each problem. The estimates should be made with upper and lower bounds and should be specified so that the motivations for the estimates can be studied.

7.5 Framework as a whole

This paragraph provides a discussion on the framework as a whole.

Validity
Based on the results from the case studies, it can be concluded that the framework fulfills the requirements that it should support the finding of usability problems and support the process of conducting a CBA of the usability effort. In both the StudyCapture and the StudyAdmin case studies a number of usability problems were found and a CBA analysis of the usability evaluation was successfully performed.
A conclusion that has been drawn based on discussions with staff members that are potential users of the framework is that they fell that it would help them in their work. In the same discussions it was argued that the results from applying the framework would be applicable in business cases at the company. Although, the future users should put the framework to test, before final conclusions on how valid it is can be made.

One goal that was set for the framework was that it should be applicable in around 40 man-hours. The exact time that it would take to use the framework has not been

measured, but based on preliminary estimates the conclusion is that using the framework in its minimal form should take between six and eight workdays. This hypothesis should also put to the test through letting the future framework users test the framework.

The prior parts of this chapter have motivated the validity for each activity that is performed within the framework but the validity of the parts in relation to each other has not been discussed. However for the framework to be valid and reliable all of the framework phases need to be present. The activities within each phase on the other hand, can be adapted to fit each evaluation and the extent of each project.

The case studies showed that the framework could be used for the evaluation of systems in use. The framework has not been tested on new systems and it has not been used for comparing systems with similar functionality. Therefore no conclusions are drawn on if the system can be used for anything else than evaluations of systems in use. Further development and testing would be needed in order for the framework to be applicable for other tasks.

Reliability

The main reliability concern with the framework comes from the limitations that have been made in time and budget for its use. In order for the framework to be executable within a timeframe that is close to 40 man-hours the results have to be based on a combination of measurements and estimations. The reliability of the results could be improved if the budget was increased so that a larger number of evaluators could be given more time. A larger budget would enable the evaluator to rely more on measures than estimations.

Since the results from the case studies have not been tested through applying changes to the systems, a conclusion about the reliability of the results cannot be drawn. Another concern is that the evaluations in the case studies were not limited to the short timeframe and the results form those can not be used as examples of how reliable the results would be when if the framework was applied with the defined time limits. If conclusions on the reliability of the framework results are to be drawn, the actual staff members that are to use it should also test the framework. Despite that the reliability has not been tested thoroughly the reliability should be high. Since the process uses both experience and knowledge from members of the system organization and usability experts. The goal is not to end up with exact measures when using the framework; it is to provide background and reliable estimates to use in a business case. The framework provides good estimates that are motivated and traceable if they are questioned, therefore the results of the framework should be reliable enough for the purpose.

Conclusions

The framework achieves the goals of guiding the process of finding usability problems and supporting a CBA of the usability effort. The framework can be executed with six to eight days but the timeframe can be increased to improve the reliability and impact of the results. The balance between the times that each activity should take can be adapted to fit each evaluation. The framework does not provide results that are 100 percent accurate but it provides estimates that are well motivated and traceable and therefore serve their purpose well.

In its present form the framework is best fitted for and should primarily be used for evaluations of systems in use. If the framework is to be used for other purposes such as evaluations of new systems and comparisons between systems, it should be tested and adapted to fit the new purpose. Staff members at the company should

further test the framework, if addition, conclusions on the validity and reliability are to be drawn. Although staff members that are potential users of the framework have already expressed that the framework will fill a function and be usable for the company.

7.6 Future development

The secondary objectives for this thesis were to investigate if the framework could be used for:

- Evaluation of systems that have not been introduced at the company
- Comparing the usability of different systems with the same purpose

As is discussed above the secondary objectives were not accomplished in this thesis. The framework in its current form should not be used for the evaluation of new systems or comparing systems. Although some parts of the framework could be used for that purpose. A suggestion for further developing the framework would be to investigate how the framework should be altered to fulfill the secondary objectives.

The main issue when evaluating new systems that the framework does not handle is the lack of users and staff with knowledge of the system that can aid the evaluation. Therefore the activities that involve those staff members would have to be altered.

The background studies phase would have to be changed and the information would have to be gathered from different people. However the information that should be gathered would be the same. The severity-rating phase would also have to be changed since system users could not set the ratings. A suggestion would be to let more evaluators rate the problems. The attitude survey would have to be taken after new users received some kind of training. The rest of the activities would also have to be altered slightly but the goals and the results should be the same.

The issue of comparing systems is much harder to tackle. This framework provides a basis for finding a number of usability problems and calculating the benefit market for them the benefit market that those usability problems receive is in no way correlated to the overall usability of the system. Therefore our conclusion is that for the comparison of new systems an entirely different approach would be needed, perhaps a second framework should be developed. Comparing systems would have to rely more on comparing tasks that the system is to fulfill. A comparison cannot be based on finding problems, more exact measurements should be found.

8 Results – The final framework

This chapter contains a description of the final proposal for a framework for usability evaluation at AstraZeneca. This proposal is based on the theory and case studies that were presented in this thesis. The framework is intended to support staff, with usability experience, in the usability evaluation process. In order for the framework to fit the organization's needs, it has been devised to support the process of prioritizing usability problems. Additionally the framework provides guidelines for performing a cost-benefit analysis of the usability effort. The framework is primarily intended to be used for evaluating and improving the usability of systems that are already in use. Many of the activities that are suggested in the framework are dependent on staff members with experience with the system and would therefore have to be substituted for other activities if the framework is to be applied to a system that is not yet introduced at the company.

The framework that is presented is based on a budget where one usability expert can spend six to eight days working on the evaluation. However the timeframe for each evaluation should be adapted to the fit the extent of each project. Suggestions on how to adapt the framework components to changing circumstances will be presented for each phase of the framework that is specified below.

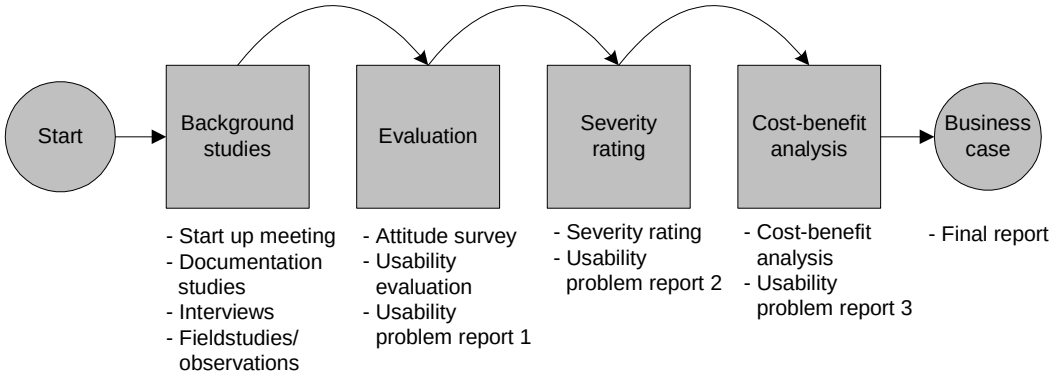

Figure 8.1 - Overview of the framework for usability evaluation

The framework consists of the following phases:

1. Background studies
2. Evaluation
3. Severity rating
4. Cost-benefit analysis

When the framework is used, the process should be documented and three problem reports should be written along the way to enable tracing of the decision that have been made and to make future evaluations easier.

Background studies
Timeframe suggestion: 2 days

The background studies are meant to form the basis for the rest of the evaluation. During this phase the evaluator is to acquire information concerning the system, its users and the context it is used in. An additional goal for this phase is to gather information on which system tasks that are the most critical and need to be evaluated in detail. This means that all necessary information that the evaluator needs for

deriving use cases have to be collected in this phase, the critical tasks should be selected based on the frequency of use, how critical it is that they are performed correctly and how problematic they have been reported to be. The evaluator can adapt the background studies, if he/she thinks that there are better-suited ways of collecting the needed information. To fit the specified timeframe, the four activities below are recommended.

Start up meeting
A meeting should be arranged with staff members that have an insight to the system that is to be evaluated and the surrounding organization. The goal for the start up meeting should be for the evaluator to establish a first overview of the system. The meeting participants should also define the purpose and goals for the evaluation, including the budget and timeframe for the project. A secondary goal is to gain the organizational knowledge needed for finding more people to contact during the later stages of the background studies. The evaluator should also be granted access to the system that is to be evaluated after the meeting.

Documentation studies
The documentation studies are to result in a better understanding of the system for the evaluator. Functional details and workflows have to be understood by the evaluator. For best use of the time given, an e-learning session, if available, is recommended. Otherwise a user's manual works as a substitute.

Interviews
This phase of the background studies is there to give the evaluator a first glance at possible usability problems and to gain a better technical understanding of the system. The evaluator should also gain an understanding of how problems are reported and questions are raised within the organization. This understanding can lead to finding additional usability problems through contacting the users involved in the process. It is good to specify the questions, which are to be used in the interview, in advance. A semi-structured way of interviewing is recommended to gain as much information as possible in as little time as possible. At least one interview with a staff member with a supporting role has to be made. This is important because the technical problems reported can help the evaluator in the evaluation.

Field study / observation
The field study / observation is an important part of the background studies. It is important because the evaluator gets a first hand look at how the system is really used and the context it is used in. Usability problems can be spotted and the user can be interviewed as the work continues. The field study / observation itself should be performed as following:

- The user should be asked to perform his usual tasks.
- The evaluator should interfere with the workflow as little as possible.
- The evaluator should ask questions if some things are unclear.
- The evaluator has to find the critical workflows for the evaluating process. These flows are critical tasks that are the most used and that are the most important not to be erroneous.

Adaptation
The activities above can be adapted to fit a larger evaluation budget. The activities that can benefit from more time given are primarily the interviews and the field study / observation. The evaluator can conduct more and more extensive interviews, but the interviews conducted have to be motivated. Several interviews with the same user type might not be necessary. The field study / observation can be performed more

than once. This gives the evaluator the chance to reflect on the findings of the first visit and develop the thoughts further.

Evaluation
Timeframe suggestion: 2,5 days

The evaluation has one main goal and that is to locate and specify as many usability problems as possible in a specified time. The input to the evaluation comes from the background studies and should be very clear to the evaluator. The critical flows that were derived in the background studies have to be specially taken care of and further evaluated in the usability evaluation. An attitude survey should also be created and distributed amongst the correct user group. The attitude survey has for purpose to evaluate the users' attitudes towards the usability in the system.

Attitude survey
The attitude survey is based on a well-known scale called system usability scale. This provides a result from where the resulting number varies from zero to 100; where 50 can be considered neutral. The survey consists of three demographic questions and then ten questions regarding the attitude, where the final results are calculated according to the following formula. First sum the score contributions from each item. Each item's score contribution will range between zero and four. For item one, three, five, seven and nine the score contribution is the scale position minus one. For items two, four, six, eight and ten, the contribution is five minus the scale position. Multiply the sum of the scores by 2.5 to obtain the overall value of system usability.

The survey should be distributed to the correct user group by the evaluator. If the wrong user group is chosen, the results might not be helpful at all. The recommended numbers of responses are at least 15 to make a proper analysis.

How the results from the survey are analyzed is up to the evaluator to decide. The demographic question helps the evaluator in deciding how to analyze the results.
- Are experienced users more satisfied?
- Does the time spent in the system per week matter?

The attitude survey can be observed in detail in Appendix A.a.

Usability evaluation
The goal of this phase is to find and specify usability problems in the system. Input to the use cases is the information that was gathered in the background studies. Two approaches are recommended in the evaluation. The first is a broad heuristic evaluation of the system based on Norman's *Design principles,* Nielsen's *eight golden rules of interface design* and Shneiderman's *Eight golden rules of interface design.* Should detect parts of the interface that do not follow the conventions for good usability. The second is a deeper evaluation of the most critical user-tasks (that were defined through the background studies) based on the same principles. The critical user tasks should also be evaluated through a flow analysis (cognitive walkthrough) with respect to if they achieved the goals that they were supposed to, in an effective way. It is important that the critical workflows are properly evaluated, because they are such an important part of the system's usability aspect

The quality of the evaluation often comes down to how much experience the evaluator has and how much time the evaluator is allowed to spend on the evaluation. The methods that are recommended are only guidelines, the evaluator is free to find the usability problems in any way that he/she wants to. As long as they

are specified so that they can be used in the later stages of the framework.

The usability problems have to be specified so that it will be easy for the evaluator to understand them later on. When the report is ready to hand over to the next phase of the project it should contain all usability problems that were found during the evaluation and background studies phases, along with sufficient explanations.

Problem report 1
The first problem report is a report where all usability problems found in the evaluation and in the background studies have to be documented. The report has to be saved for future considerations, If it is decided that the evaluation gets a larger budget than expected, it can be of great help to have this problem report, so that problems, that were not further evaluated, quickly can be investigated. How the specific problems are documented is up to the evaluator, but it is necessary for the evaluator to write the problems so that another evaluator can understand them later on.

Adaptation
If more time is to spend in the evaluation phase of the project, a good way of increasing the quality is to extend the number of evaluators. The validity for one evaluator is much lower than if you use more. More time used by one evaluator is also a way of increasing the quality of the resulting usability problems. If more time is spent, the evaluator can make deeper analyses and get better results. A third way of increasing quality is to analyze more workflows. This means that more time has to be spent, but better system coverage is achieved.

Usability severity rating
Timeframe suggestion: 1 day

The usability severity-rating phase of the framework is there to prioritize and select the most important problems that have been found in the usability evaluation. Since the time that is spent on a usability evaluation is limited, it is of high importance that there is a way to prioritize which usability problems that should be further evaluated and investigated. The input to this phase is usability problem 1 from the evaluation phase. The results from the severity-rating phase will be used to prioritize the usability problems according to severity. The severity rating will also be used as input to the cost-benefit analysis.

Rating
To enable a prioritization of the usability problems each problem has to be rated according to the frequency at which it's consequences appear and impact that the consequences have. The first activity that the evaluator performs in this phase is a selection of the problems that he/she considers to be the most critical in problem report 1. The selected problems should then be specified in a new problem report so that others can understand them.

After the problem specification the evaluator should distribute the problem report along with a survey to selected staff member who have experience with the system. The staff members that estimate the severity ratings should be selected from primary user groups and other system stakeholders such as super users, system owners and administrators. In order for the results to be reliable there should be at least five people rating the problems. The evaluator and staff members that have been a part of the background study and reported problems should not be part of this activity.

For each problem in the report there should be three corresponding questions in the survey. The web-survey that is recommended can be read in Appendix A.b. The results from the survey should be summarized through:

- Calculating an average number for each question (0-4).
- Multiplying the average numbers for the three questions to receive a severity rating on the problem.

The severity ratings can then be used to prioritize the usability problems.

Problem report 2
The results from the severity-rating phase should finally be documented in a second usability problem report along with the problem descriptions that were distributed along with the severity survey.

Adaptation
The severity rating activities can be adapted and improved to fit a larger evaluation budget. The reliability can be improved through adding more problems for rating and through letting a higher number of key staff members rate the problems.

Cost-benefit analysis
Timeframe suggestion: 1 day

The cost-benefit analysis (CBA) has the purpose of calculating the costs and the benefit market associated with each usability problem. The CBA can support decisions concerning which usability problems that should be fixed, as well as to motivate the usability effort towards the rest of the organization and the system developers. The usability problems that have been found to have the highest severity rating should be chosen for analysis.

Cost-benefit analysis
The cost-benefit analysis should be conducted through a meeting with key staff members. The input to the meeting should be:

- Usability problem report 2 – before the meeting the evaluator should make a selection of the problems that have received the highest severity rating and bring those to the meeting.
- The number of user that are affected by each problem.
- The number of new users that are introduced to the system every year.
- The cost for one full time equivalent (FTE).
- The amount of time that is considered to be learning time for a new user.
- The possible consequences of each problem according to the evaluator.

The evaluator should lead the meeting and provide usability expertise, in addition to that two or more experts on the system with knowledge of the organization, system details and system task should participate. This combination of people should be able to judge each problem according to frequency and the time spent if it occurs.

The first activity during the meeting is to analyze each usability problem and predict the possible consequences if it occurs. The consequences can with advantage be described in a figure.

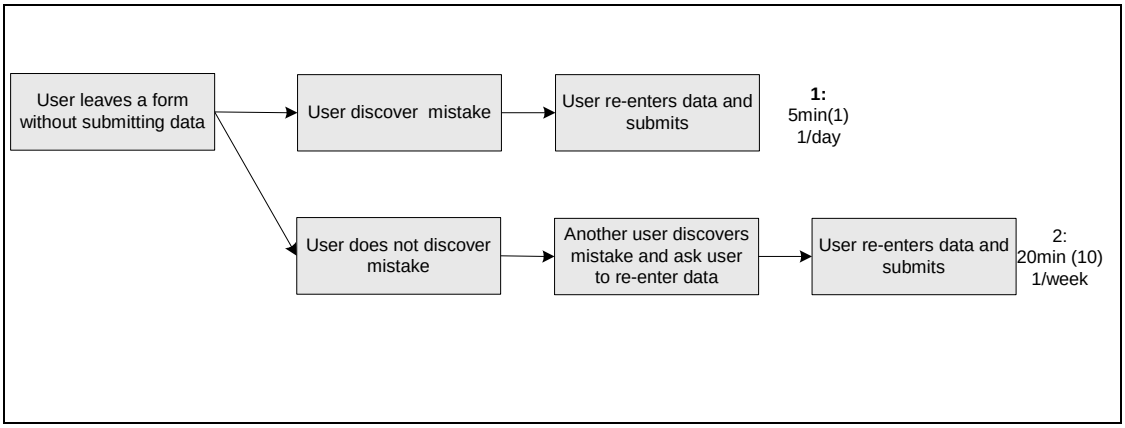

Figure 8.2 - Example of an analysis of the consequences of a usability problem in the StudyCapture case study

The next step is to estimate the frequency of each consequence and the time it would consume if it occurred. To make the estimates more reliable upper and lower bounds should be defined. The estimates should be made based on the frequency and impact that a number of staff members predicted during the severity rating activity. The possible savings in learning time per new user should also be estimated based on the time that is considered to be learning time and the severity of the problem.

Based on the estimates of the consequences and saved learning time calculations can then be made on what the benefit market for each problem is. The calculations are made through multiplying the savings per user with the number of users that are affected by the problem.
The following formula can be used for the calculations of the benefit market of a problem:

*(problem min/day/user * nbr of users) / (total work minutes/day) = number of FTEs*

To calculate the benefit market for the saved learning time the following formula can be used:

*Possible savings in time per new user * number of new user per year*

The timesaving can be calculated into monetary value through multiplying with the cost for a FTE. It is important to remember that the calculations result in the benefit market for fixing each problem, meaning the savings that could be made if the usability problems were fixed to 100 percent. This is an impossible result to reach for most problems and therefore there has got to be a suggestion on how the problem is to be fixed if more detailed calculations are to be made. The suggested solutions can then be analyzed further and a final calculation of the cost-benefit for the usability evaluation can be calculated according to table 8.1.

Table 8.1 - Summarized CBA calculation

Total cost reduction (+)	
Costs for fixing the problems (-)	
Evaluation costs (-)	
Total benefit (=)	

Although it is important to remember that this summary is not supposed to be made when using the framework. It does not fit into the timeframe; the continued analysis is

a follow-up of the first evaluation. The CBA within the framework execution should stop at the calculation of the benefit market for each problem.

Problem report 3
After the CBA meeting the results have to be specified. A third problem report should be specified. It should contain a description of the problems that were analyzed during the meeting. It should also specify the results of the severity rating for each problem along with the analysis of the problem consequence. The results from the CBA analysis should also be documented. Finally the report should contain a summary of the CBA of all the usability problems along with the results from the attitude survey.

Adaptation
The CBA can be adapted to fit a more extensive evaluation or increase the reliability of the results through analyzing more usability problems, having more stakeholders make the estimations, or following up the estimations.

Documentation
Timeframe suggestion: 1 day

A parallel process to the evaluation process itself is the documentation process. It is important to document all input and output between the different stages of the evaluation process in order to keep track of the progress.

The final report from the project should document all activities that were performed. The results from the evaluation must also be documented along with the three problem reports. It is important not to spend too much time on writing the report; instead the focus should be on the other activities since there is a strict time frame.

9 Conclusions

The main question of this thesis is; *is it possible to devise a general framework that guides AstraZenecas R&D staff in the usability evaluation process*? The question has been answered through the development of a framework for usability evaluation that is to be used by staff at AstraZeneca.

The framework has been tested through case studies that were conducted at the company. Two separate case studies were conducted and both of them showed that the framework supports the process of finding usability problems in IT-systems. The results also showed that the framework could be used for conducting a cost-benefit analysis of the usability effort. The staff at AstraZeneca that is intended to use it has not tested the framework. Although, the framework has been developed in close cooperation with the staff and the results have been accepted by them. This shows that the framework not only serves the purpose, but also fits the setting it is to be used in. Another goal for the framework was that the timeframe for applying it should be around 40 man-hours in order for the framework to be practically applicable in the organization. Additionally the framework should fit into the organization through providing input to business cases. The case studies indicated that the execution of the framework in its minimal form should take around six to eight workdays, which exceeds the time limit by a small amount of time. However it was shown that in order for the evaluation result to be valid and reliable enough for the use in business cases, the slightly longer timeframe was motivated.

The framework was developed based on a compilation of theories that were chosen from some of the most prominent authors within the usability field. This strong basis in current research supports assumptions that have been made in the thesis and should ensure that the framework is theoretically accepted. The theoretical background also helps keep the framework general so that it would be possible to use it in another setting with slight alterations.

The framework is best suited for usability evaluations of systems in use. Many of the activities in the framework are highly dependent on users and staff with knowledge of the system in evaluation. Thus it can be concluded that the framework in its present form is not well suited for evaluating systems that have not been introduced or comparing systems with similar functionality. Some of the parts of the framework could suit those purposes, but the framework as a whole would have to be further developed and adapted to fit those tasks.

The qualities of the results that come out from using the framework are highly dependent on the user of the framework and the budget that the evaluation is given. The user needs to possess knowledge on usability in order for the framework to be usable and a more experienced usability expert would most likely lead to better results. The budget that is given to the evaluation affects how much time that can be spent on each activity and therefore impacts the quality of the results. The final results are based on a series of estimations and are therefore impacted by both the knowledge and experience of the evaluator and the selection of staff members that aid in the evaluation and participate in the activities of the framework. Since the results of the framework are dependent of estimations an important feature is that all results can be traced and motivated. The framework provides guidelines for specification that enables traceability of the estimations.

Despite the uncertainties concerning the framework, it can be concluded that its potential for improving the usability effort at AstraZeneca is high. The framework has

proved to meet the main goals that were set up for it, and it has been well accepted by staff members at company. The framework can help AstraZeneca improve the usability of their IT-systems and thereby save them time and resources.

10 References

Literature

Balic, M., Ottersten, I. 2004. *Effektstyrning av IT – Nyttan uppstår I användningen.* Liber Ekonomi.

Berndtsson, J., Ottersten, I. 2002. *Användbarhet i praktiken.* Studentlitteratur.

Ekholm, M., Fransson, A. 1992. *Praktisk intervjuteknik.* Norstedts.

Jönsson, B. 2005. *Människonära design.* Studentlitteratur.

Mattsson, B. 2004. *Kostnads-nyttoanalys.* Räddningsverket.

Nielsen, J. 1999. *Designing web usability.* New riders publishing.

Nielsen, J. 2006-11-16. *Severity Ratings for Usability Problems*. http://www.useit.com/papers/heuristic/severityrating.html

Nielsen, J., Mack, R. 1994. *Usability inspection methods.* John Wiley & Sons, Inc.

Nielsen, J., Tahir, M. 2001. *Homepage usability, 50 websites deconstructed.* New riders publishing.

Norman, D. 1998. *The Design of Everyday Things*. The MIT Press.

Norman, D. 1993. *Things that make us smart.* Perseus books.

Preece, J., Rogers, Y. Sharp, H. 2002. *Interaction design – beyond human-computer interaction*. John Wiley & Sons Inc.

Rubin, J. 1994. *Handbook of usability testing.* John Wiley & Sons, Inc.

Schaffer, E. 2004. *Institutionalization of usability.* Addison-Wesley Computing.
Shneiderman, B., Plaisant, C. 2004. *Designing the user interface.* Addison-Wesley Computing.

Vagelin, A. 2005. *Usability evaluation of a system in use.* Master thesis.

Walla, E. 1990 *Så skriver du bättre tekniska rapporter.* Studentlitteratur.

Electronic sources

Brooke, J. 1986. *SUS - A quick and dirty usability scale.* http://www.cs.concordia.ca/~soen357_4/SUS-paper.pdf.

Course Material for the course Human-Computer Interaction. LTH. 2004-12-05. http://wwwold.eat.lth.se/Kurs/Material/MAM061/.

ISO. 1998. *ISO DIS 9241-11.* http://www.iso.org.

Nielsen, J. 1994. *Guerrilla HCI: Using Discount Usability Engineering to Penetrate the Intimidation Barrier.* http://www.useit.com/papers/guerrilla_hci.html.

University of Texas. 2006-09-27. *Shneiderman's "Eight Golden Rules of Interface Design".* http://www.cs.utexas.edu/users/almstrum/cs370/elvisino/rules.html.

UsabilityNet. 2006-11-10. *Human centered design processes for interactive systems*. http://www.usabilitynet.org/tools/13407stds.htm.

Appendix A - Surveys

a Attitude survey

The attitude survey that was used is an interpretation of the SUS survey *(*Brooke, 1986). It was extended with the first three questions and in some places changes have been made to adapt the survey to the framework.

The SUS scores have a range of zero to 100. To calculate the SUS score, first sum the score contributions from each item. Each item's score contribution will range between zero and four. For item one, three, five, seven and nine the score contribution is the scale position minus one. For items two, four, six, eight and ten, the contribution is five minus the scale position. Multiply the sum of the scores by 2.5 to obtain the overall value of SU.

Additional questions (Example questions)

1. What is your main function when working with the system?

2. How long have you had this role? And, have you had any other tasks, in the past, involving this system?

3. Approximately, how many hours a week do you use the system?

Authors' interpretation of SUS - *A quick and dirty usability scale (*Brooke, 1986).

1. I find the system well worth using.

1 = I do not agree, 5 = I agree completely

2. I find the system unnecessarily complex.

1 = I do not agree, 5 = I agree completely

3. I think the system is easy to use.

1 = I do not agree, 5 = I agree completely

4. I think that I would need the support of a technical person in my continuous work.

1 = I do not agree, 5 = I agree completely

5. I find the various functions in this system are well integrated.

1 = I do not agree, 5 = I agree completely

6. I think that there is too much inconsistency in this system.

1 = I do not agree, 5 = I agree completely

7. I imagine that most people have learnt this system very quickly.

1 = I do not agree, 5 = I agree completely

8. I find the system very troublesome to use.

1 = I do not agree, 5 = I agree completely

9. I felt very comfortable using the system.

1 = I do not agree, 5 = I agree completely

10. I needed to learn a lot of things before I could get going.

1 = I do not agree, 5 = I agree completely

b Usability severity ratings

This survey is used to rate problems in the usability problem report. Each usability problem corresponds to a set of three questions. An example the set of questions for a usability problem is shown below.

Usability problem nr. X

Problem frequency

1. How frequent do you think that the user group it applies to uses the functionality, which is affected by the usability problem?

- ☐ Yearly
- ☐ Monthly
- ☐ Weekly
- ☉ Daily
- ☐ Several times a day

2. Out of the times that the functionality is used, how often do you think that this usability problem affects the user in his/her work with the system?

- ☐ Never
- ☐ At rare occasions
- ☉ Every now and then
- ☐ Often
- ☐ Every time

Problem impact

3. How serious do you think that the possible consequences of this usability problem are?

- ☐ Not an issue
- ☉ Minor
- ☐ Serious
- ☐ Major
- ☐ Critical