

AALBORG UNIVERSITY

- IN COLLABORATION WITH NOVO NORDISK A/S

MASTER THESIS

ENGINEERING PSYCHOLOGY

Alternatives to Sampling of Hard-to-Reach Populations

Designing a Method for Identifying Substitute Samples

Author

SIMON KRING NIELSEN

June 6, 2018





AALBORG UNIVERSITY
STUDENT REPORT

Department of Electronic Systems

Frederiks Bajers Vej 7
DK-9220 Aalborg Ø
<http://es.aau.dk>

Title:

Alternatives to Sampling of Hard-to-Reach Populations: Designing a Method for Identifying Substitute Samples

Semester:

10th semester, Master Thesis

Semester theme:

Engineering Psychology

Project period:

Spring 2018, February 1st to June 7th 2018

Supervisor:

Associate Professor, Lars Bo Larsen

Corporate Supervisor:

Rasmus Klinck

Project group:

PDP10-1083. 18gr1083

Simon K Nielson

Simon Kring Nielsen

Abstract:

Conducting evaluations requires users from the target population, which in rare occasions is near impossible to obtain. Through the project it is investigated how these situations can be approached, without relying on the existing methods of hard-to-reach samples. The investigation of sampling refers to reviews of literature and previous research, but also entails the collection of empirical data on the subject of sampling.

Through research a method is developed with the goal of enabling identification of appropriate use of substitute samples in evaluations. The method is further validated through a workshop conducted with Novo Nordisk. The validation produced suggested changes and optimizations of the method, in order for it to function in everyday work. The changes mainly suggested were to increase agility of the method.

In conclusion the method fulfills the intended goal, while there is still room for improvements.

Copies: 2

Pages: 221

Date of Completion: June 7th
2018

Danish Summary

Rapporten tager udgangspunkt i sampling og de problemer der eksisterer når en population er utilgængelig. Derudover ønskes det at opnå en metode til identificering af substituerende samples. Initierende opstilles tre spørgsmål som ønskes besvaret. Disse er:

1. Hvordan kan sampling gribes an når populationen man ønsker at teste på ikke er tilgængelig?
2. Hvordan kan sampling gribes an når konfidentialiteten af et produkt overskrider hvad der er fornuftigt at eksponere uden for virksomhedsrammerne?
3. Hvilken effekt har kvalitative og kvantitative tilgange på sampling?

I besvarelsen af disse tre spørgsmål bliver litteraturen omkring sampling gennemgået. I gennemgangen af eksisterende litteratur findes det at metoderne i høj grad fokuserer på at en sample skal være repræsentativ og taget fra den pågældende population. De eksisterende metoder, der findes til at opnå adgang til populationer, der er svært tilgængelige findes dog ikke specielt veldokumenterede. Derudover synes metoderne at benytte sig af ukontrollerbare variable, såsom at afhænge af deltageres sociale netværk indenfor den pågældende population.

Da gennemgangen af litteraturen ikke bidrager til en tilfredstillende besvarelse af spørgsmålene opstilles et interview. Formålet med interviewet er at klarlægge hvordan sampling foregår i virksomhedsøjemed. Tina Øvad og Morten Purup giver gennem interviews deres perspektiv på sampling, som i høj grad hjælper til at besvare spørgsmålene. Udeover denne empiriske data gennemgås også tidligere research i håbet om at identificere studier med lignende problemer.

Opsamlingen af disse empiriske data muliggør besvarelse af to ud af de tre spørgsmål. Når konfidentialiteten af et produkt overskrider hvad der er fornuftigt at eksponere uden for virksomhedsrammerne, eksisterer to valgmuligheder. Ét er at teste det internt og afhænge af ekspert evalueringer. Et andet er at teste det på brugere alligevel og lægge sin tro i at en NDA er fyldestgørende nok. Samtidig er problemet egentlig selvskabt. Hvis man vælger ikke at eksponerer sit produkt for brugerne, så tilvælger man også mindre valide tests.

Med henblik på kvalitative og kvantitative tilganges indflydelse på sampling, så synes denne ikke at være specielt stor. Der er forskel mellem de to tilgange, men generelt kan det siges at de forholdsregler, der eksisterer inden for de to tilgange også gør sig gældende for en eventuel substitueret sample.

Spørgsmålet om hvordan sampling gribes an når populationen man ønsker at teste på ikke er tilgængelig blev udelukkende uddybet, gennem litteratur og empiri. Dog forekom forslag til hvordan en metode kan udvikles til at omgå problemet. Formålet med metoden er at identificere hvilke forudsætninger populationen har, som adskiller sig fra andre popula-

tioner. Dernæst er ønsket at identificere hvilke forudsætninger der er påkrævet i den specifikke evaluering. Endeligt ønskes det at sammenligne forudsætningerne. Til at opnå disse tre kriterier benyttes 1) personaer, 2) Hierarkisk Task Analyse og 3) Systematisk Menneskelig Fejl Reducering og Forudsigelse (SHERPA).

Metoden var initialt tænkt at skulle verificeres ved at benytte de opstillede metoder til at udpege en substituerende sample og udføre et studie dermed. Studiet var en replikation af et tidligere studie, udført på diabetikere, hvilket ville tillade en sammenligning af den substituerende sample og samplen fra den pågældende population. Denne tilgang blev forsøgt, men blev umuliggjort af manglende stimuli.

Dette udledte at metoden i stedet blev verificeret gennem en workshop. Fokus blev flyttet fra at validerer metoden resultatlæssigt til at validerer metoden brugs- og funktionaltetsmæssigt. Gennem workshoppen blev områder med mulighed for optimering identificeret og derudover blev det påvist hvordan metoden kunne bruges. Endeligt blev metoden brugt af forfatteren selv, grundet manglende detaljer fra workshoppen. Metoden kan inden for flere områder optimeres, men grundet ressourcerne for projektet, har det ikke være muligt at validerer de potentielle optimeringer. Heraf skal det forstås, at de optimeringer der blev fundet frem til, udelukkende er spekulationer, indtil det er muligt at verificere dem.

Det konkluderes at metoden og fremgangsmåden kan afhjælpe problemet omkring ikke-tilgængelige populationer i evalueringer. Det gøres ved at inddrage populations forudsætninger, de forudsætninger evaluerings mål påkræver samt at klarlægge hvordan disse forudsætninger til sammen kan identificere og danne de situationer hvor det er muligt at bruge en substituerende sample. Gennem metoden muliggøres det også at fastsætte de nye kriterier der gør sig gældende for den substituerende sample.

Preface

This thesis written solely by the Author. The author is from Engineering Psychology student, 10th semester at the Institute of Electronic System at Aalborg University. The period in which the project took place is from February 1st to June 7th 2018. Lars Bo Larsen, Associate Professor at the Institute of Electronic Systems has been supervising the project and provided ongoing input throughout the period of the project.

The theme of the thesis is "Sampling" and through the report it is ought to obtain an alternative to the existing methods of sampling hard-to-reach populations. The focus is especially to investigate the possibility of substituting the hard-to-reach target population with other more accessible populations.

The thesis uses the Harvard Reference Method. This implies all references consisting of author name(s) and year of publication. A full list of references for the literature used throughout the project can be found at the end of the report on page 95. In addition to references a list of figures and a list of tables i provided. Further, nine appendices have been attached and starts on page 100. The appendices are developed by the author and functions as support for some of the argumentation provided in the report. When referring to a specific appendix in the text, an identifier for specific appendix is given in coherence with a page number, e.g., "Appendix A on page 111".

Acknowledgement

In the collaboration with Novo Nordisk, multiple people have reached out and helped in situations where the author wouldn't have been able to accomplish. The employees in *Department 5072, User Research, Usability and User Communication* are therefore further acknowledge. The acknowledgement is due to helping in the situation of identifying studies and investigating whether these could be used as a part of the thesis.

Besides, the author would like to acknowledge Tina Øvad and Morten Purup for participating in interviews and provide their perspectives on sampling. The aspects covered took major part in the writing of the thesis. Lastly, the author thank everyone who have participated in the workshop or generally just given input to the development of method.

Contents

	Page
DANISH SUMMARY	ii
PREFACE	iv
1 INTRO	1
1.1 THE GOAL OF USABILITY AND USER EXPERIENCE RESEARCH	1
1.2 WHEN REAL USERS ARE NOT AVAILABLE	2
2 SAMPLING IN GENERAL	4
2.1 THE DEFINITION OF A SAMPLE	4
2.2 WHAT IS A GOOD SAMPLE?	5
2.3 VERIFYING THE SAMPLE	6
2.4 DIFFERENCES IN QUALITATIVE AND QUANTITATIVE APPROACHES	9
2.5 THE SELECTION OF A SAMPLING METHOD	11
2.6 SUMMARY OF SAMPLING	13
3 EMPIRICAL DATA ON SAMPLING	14
3.1 THE VALUE OF CONTEXT	15
3.2 THE INFLUENCE OF LOCATION, PARTICIPANTS, AND RESOURCES	16
3.3 ALTERNATIVES TO USER INCLUSION	19
3.4 INTERVIEWS WITH B&O AND TECHNOLUTION	20
3.5 SENIOR UX & USABILITY LEAD, TINA ØVAD	21
3.6 SENIOR DESIGN ENGINEER & HUMAN FACTORS SPECIALIST, MORTEN PURUP . .	25
3.7 DISCUSSION OF LITERATURE, THEORIES, AND EMPIRI	28
4 HOW TO APPROACH SAMPLING	31
4.1 ANSWERS TO INTERIM RESEARCH QUESTIONS	31
4.2 HOW TO PROCEED FROM HERE?	34
5 DESIGNING THE METHOD FOR SUBSTITUTE SAMPLING	35
5.1 ANALYSIS OF THE TARGET POPULATION	36
5.2 ANALYZING THE EVALUATION THROUGH HIERARCHICAL TASK ANALYSIS	39
5.3 COMPARING THE TARGET POPULATION AND TASK ANALYSIS	42
5.4 THE FULL METHOD	46
6 VALIDATING THE METHOD	48
6.1 DESIGN OF WORKSHOP	49
6.2 DATA COLLECTION METHOD	54

7 CONDUCTION AND REVIEW OF THE WORKSHOP	55
7.1 THE WORKSHOP IN GENERAL	56
7.2 CREATING A FOUNDATION OF PERSONAS	57
7.3 USING TASK ANALYSIS ON A CASE	60
7.4 PREDICTING HUMAN ERRORS WITH SHERPA	66
7.5 COMPLETE USE OF THE METHOD	68
7.6 SUMMARY OF FINDINGS AND SUGGESTED CHANGES	81
8 DISCUSSION OF SAMPLING	85
8.1 LITERATURE, PREVIOUS STUDIES, AND THE RESULTS	85
8.2 THE METHOD AND FURTHER VERIFICATION	86
8.3 RELATED ISSUES	87
8.4 MAINTAINING CONTEXT AND MOTIVATIONS	88
8.5 FUTURE WORK AND NEEDS	88
9 CONCLUDING ON SUBSTITUTE SAMPLING	90
REFERENCES	92
LIST OF FIGURES	96
LIST OF TABLES	98
APPENDICES	100
A NOVO NORDISK AND SAMPLING DIFFICULTIES	101
B ASSESSMENT OF METHODS FOR QUALITATIVE ANALYSIS	102
C QUANTITATIVE AND QUALITATIVE APPROACHES	111
D TRANSCRIPTS OF INTERVIEW	118
E ANALYSIS OF INTERVIEW	139
F DETAILED DESCRIPTION OF METHODS	160
G DIABETES IN GENERAL	174
H THE WORKSHOP AND THE TOOLS FOR CONDUCTION	180
I THE KJ-TECHNIQUE	219

Chapter 1

Intro

1.1 The Goal of Usability and User Experience Research

Usability testing and user experience research share one common goal; to become further enlightened within a given area of interest. The characteristics distinguishing the two evaluation types are that usability is product related and user experience research is person related. Usability and user experience research are however very difficult to separate. Good usability of a product can lead to good user experience, and good user experience can sometimes offset participants perception of how good the functionality is. Usability and user experience are coherent and not two independent areas in user evaluations (Moczarny et al., 2012). Usability testing has the goal of identifying problems within the use of the product (Dumas and Redish, 1999, p. 22). These problems can be anything from cosmetic to critical, where a critical problem often can be described as a problem hindering the intended use. The overall goal of usability testing is to optimize and improve the usability of a product for the intended use. For usability testing Dumas and Redish (1999) implies that all usability testing shares these five characteristics:

- A primary goal to improve the usability of a product. Each iteration develops specific goals and concerns. These are incorporated into the planning of the tests.
- The participants represent real users when interacting with the product.
- The participants do real tasks when interacting with the product.
- One observes and records what the participants do and say.
- One analyzes the data, diagnose the real problems and recommends changes to fix those problems.

Oppositely, user experience research focuses on the experience a person obtain when interacting with the product. The experience a user has includes multiple subjective aspects. These subjective aspects could be, feelings, motivation, former experiences with the product domain. Subjectivity makes user experience research harder to generalize, as peoples attitude, motivation and ideologies usually are very different. The individual differences have a significant impact, and it can be challenging to establish shared characteristics across all user experience research. An argument for this is that user experience is not exactly measurable, in the sense that user experience changes, depending on the type of person, context, and product.

A Sample of Real Users

It is common knowledge that people in a sample should be representative for the end-users. This knowledge applies to both usability testing and user experience research. The requirement of using participants representing real users in studies originates from the idea that end-users who are going to live with the product will have different subjective opinions and possible prioritize differently than others. Furthermore, end-users have higher knowledge and experience in the product domain being tested. Participants with more experience than end users can result in unidentified usability problems and users with less experience can result in proposed changes that are not improvements for end users (Dumas and Redish, 1999, p. 23).

Additionally Rubin and Chisnell (2008) states that users participating should have backgrounds and abilities equal to those from the population being targeted to be representative. He also states that the test results only will be valid if the sample used in the studies are typical users of the product. He does, however, mention that the sample also could consist of users being as close as possible to the target population. 'Wrong' users should however never appear in a sample. Wrong users would be a sample of people who do not share any characteristics with the end-users. Using what Rubin and Chisnell (2008) refers to as 'wrong' users will undoubtedly make results questionable and of limited value, no matter the amount of effort (Rubin and Chisnell, 2008, p. 115).

1.2 When Real Users Are not Available

In agreement with both Rubin and Chisnell (2008); Dumas and Redish (1999) most if not all researchers state that drawing samples from the target population will yield the best and most valid results. Being interested and focusing on samples, is typically due to a wish of reaching the users who are targeted to use products. The reasoning for reaching these users can variate highly, but often the wish is to obtain insights from these users that would not be possible to obtain otherwise. The product should also be evaluated and verified by these users as they are a part of the potential buyers of the product. In the literature the group of users that one is reaching out to is defined as the *target population* and the users who are recruited and talked to are referred to as the *sample*.

Even though a joint agreement on the subject of sampling exists, inevitably there are situations where sampling users from the target population will not be possible. The situations where a matching sampling is not obtainable could be due to limited resources or hard-to-reach populations. Furthermore, it could be that the level of confidentiality exceeds what is reasonable to expose outside a business framework or it could be a combination of them all.

Not exposing a product complicates how to identify and sample from a population that somehow represents the actual target population that on beforehand is known not to be available. In addition to these limitations, there is also the statistical treatment of data. By literature, a sample is only random, when it is picked randomly from the target population. As the population often consist of multiple cultures, it can be hard to obtain a truly random sample. Most of the time a truly random sample is not obtainable, as all samples are drawn with some convenience., e.g., all participants in the sample being from the same country. Therefore, it seems that some of the presumptions about sampling existing in the literature affect how sampling should be approached.

A high-level description of the complications is to identify a new population that also represents the target population that is known to be unavailable. In a more detailed manner, it is to identify the characteristics of the actual users and afterward investigate which of these characteristics that have to be the same in an evaluation. Depending on the situation, some characteristics can maybe be adjusted while others can be manipulated or even eliminated. From a sample point of view, the characteristics needed can be anything. Therefore, only characteristics influencing the sample are in scope. Influencing characteristics can, however, be hard to define. It could be imagined that physical and psychological aspects are the bare minimum needed for understanding the identity of the target population. Such characteristics could be used for finding a new population where a sample can be drawn to 'replicate' and 'substitute'.

Definition of Problem

Several issues of sampling hard-to-reach population have been stated. The scope of the thesis is to answer and find solutions to these issues, but not limited, to investigate these with the purpose of identifying potential solutions. Initially, the scope of the thesis is to answer the below problem statement and interim research questions. The literature of sampling is revisited to acquire additional knowledge on the subject, which can help answer the questions. The literature will give insights to the issues and can potentially contribute as argumentation to the answering of the questions.

Problem Statement

When it is not possible to sample from a target population, which sampling-techniques are then applicable to gather additional knowledge on a given subject?"

Research Questions

1. How is sampling approached, when the target population is not possible to withdraw a sample from?
2. How is sampling approached, when the confidentiality of a product exceeds what is reasonable to expose outside a business framework?
3. What effect do qualitative and quantitative approaches have on sampling?

Chapter 2

Sampling in General

Answering the now established research questions requires a more profound knowledge of the subject of sampling. An examination of literature and the theories on sampling is necessary because sampling differentiates highly based on the method and domain of application.

A consensus on sampling does exist but in many cases sampling is also considered a complicated issue. These issues regard how sampling should be approached on different occasions, as sampling is defined to be very situational and to be decided based on the experimental design. Experimental designs can variate highly, and many variations of sampling are therefore present in the literature. The many variations also cause overlaps between the methods. Some of the differences and overlaps are due to the quantitative and qualitative approaches. A review of both quantitative and qualitative approaches to sampling is done, as there is a potential for drawing parallels between the two. While the following sections detailedly cover theory on sampling, the main findings are summarized at the end of the chapter, to recap the essentials.

2.1 The Definition of a Sample

When conducting research, it is for the sake of knowledge. No matter if it is an answer to a general research question within a specific scientific topic or if it is an exploratory genetic study trying to map complex human behavior. It is, however, no matter subject, often sought to generalize the findings, where the best way of doing so is to obtain data from all instances of the area of interest (Field and Hole, 2003, p. 110). Collecting data from all instances can, however, be troublesome to obtain, as this requires resources and time that typically isn't available. Obtaining data from the whole target population requires data on every single instance, e.g., products for the treatment of diabetes mellitus type 2 would require data from every type 2 diabetics worldwide. Aforementioned, this is not possible due to time and resources, and instead, data should be collected from a small subset of the population which is known as the *sample*. A well-taken sample of the population should provide data that sufficiently represent the general opinions and behaviors of the whole population, thus allowing for a generalization without using the whole population (Field and Hole, 2003, pp. 110-111).

2.2 What is a Good Sample?

A good sample can be described in many ways, as 'good' can be seen from many different points of view. Already stated, a good sample will most likely be representative of a target population, hence making it possible to generalize findings and obtaining insights from credible sources. 'Good' can however also relate to the convenience of the sample drawn from the population, i.e., a convenience aspect could be the sample not being too demographically spread or having filled in all slots of the original schedule making the sample fit the researchers time. Demographic and scheduling factors are very dependent on the number of resources available and not the characteristics of the population. As the convenience aspect neglects the user characteristics, the demographic and other high-level criteria might be categorized as recruiting criteria and not sampling criteria.

In short Kothari (2004) list five characteristics a sample should fulfill to be categorized as a 'good' sample.

- (a) The sample must result in a truly representative sample.
- (b) The sample must be such which results in a small sampling error.
- (c) The sample must be viable in the context of funds available for the research study.
- (d) The sample must be such so that systematic bias can be controlled in a better way.
- (e) The sample should be such that the results of the sample study can be applied, in general, for the population with a reasonable level of confidence (at least for quantitative).

Evaluating a samples quality is also possible, by investigating the quality of answers provided by the sample. In quantitative research, some participants might excel in all task given and can afterward put words on why they interacted, in a certain way. Other might seem horrible in their handling and are not able to explain why they do not understand how to use the product. The same applies to qualitative research, where some participants seem 'better' to pass on information and interpretations about the product (Marshall, 1996, p. 523). Users not being able to communicate why the product is hard to use, is very frustrating, because this data is beneficial for further development. A parallel can be drawn to the statement, that users do not know what they want, but they really know what they do not want. Here the valuable data is also in the unanswered question.

One way of improving the chance of reaching a good sample is by knowing the users. Concerning the problem statement, knowledge of the population cannot be disregarded. The information is too valuable. Typically some criteria for the sample is set, but if these prove challenging to recruit from compromises have to be considered. In most cases, compromises lead to a less strict set of screening criteria. Knowing the users could provide vital information to which aspects to adjust to enable recruitment of participants who are representative of the target population.

Know the Users

Knowing the users one are trying to design for is always essential for developing a successful product, as this helps to identify and accommodate user needs. Muller et al. (2001) expressed that popular advice for reaching a matching sample was: "*Know thy users*". Knowing the target population helps to identify where a sample should be drawn when evaluating a product. The knowledge is also useful if the target population is unavailable. The knowledge can be used for identifying other populations suitable for the evaluation. Explicitly greater knowledge of the target population entails a higher chance of reaching a matching and representative sample. It is, however, hard to pinpoint what kind of knowledge that is supporting the process of sampling. It can be preconceived that the target populations perceptual- and cognitive skill set in coherence with the essential user attributes together will be able to create a general identity of the users in the target population.

Knowing the identity of the population involves knowledge of motivations, needs, attitudes and the skill set affiliated with the specific population. Being able to identify these aspects of the target population also enables the identification of other population to sample from, if the target population is unavailable. Ultimately having argumentation for populations being alike enables the researcher to generalize results to all populations deemed similar.

Literature adds additional requirements for a true representative sample. While the sample might contain representative users, the overall sample would also have to match the target population. For it to do so, the distribution of users in the sample should also match the distribution from the target population, e.g., if ten percent of all people with diabetes were Afro-Americans, the sample should also contain this (Muller et al., 2001, p. 1). To the author's knowledge, considering this criterion is not done in much research. Obtaining such a construct of a sample, the variables impact also have to be deduced. The variables' impact on the representativeness of the sample has to be identified. There is no reason for having ten percent of Afro-Americans if it can be proved that the cultural or racial difference will not have an impact in the study.

2.3 Verifying The Sample

If a sample is taken from a target population and adjusted according to the distributions of, i.e., race, one might be interested in checking if the data collected throughout the research can be called representative. In this case, it is important to remember, that the ultimate goal of sampling is to enable generalization of the findings to the entire target population.

Distribution of Data

To verify if the sample is representative and generalizable different approaches can be applied. An investigation of data distribution can be applied to quantitative data. Investigating the distribution can tell if the data is normally distributed or not. The distribution will result in a mean and a standard deviation. The standard deviation is an expression of accuracy. The lower the standard deviation is, the higher is the accuracy.

Investigating the distribution will show how the answers of each participant are, compared to the other participants in the given sample. It will not verify how well the sample matches the target population. Verifying a sample match would require multiple evaluations with different samples. One sample would have to be from the target population since the verification is a comparison of the distribution (mean and standard deviation). Figure 2.1a illustrates the process, showing that mean will differentiate between samples. If means varies between samples, this would most likely also be reflected in the standard deviation of each sample. (Field and Hole, 2003, p. 132-133). In addition to the illustration, figure 2.1b is a plot showing the frequency distribution of sample means.

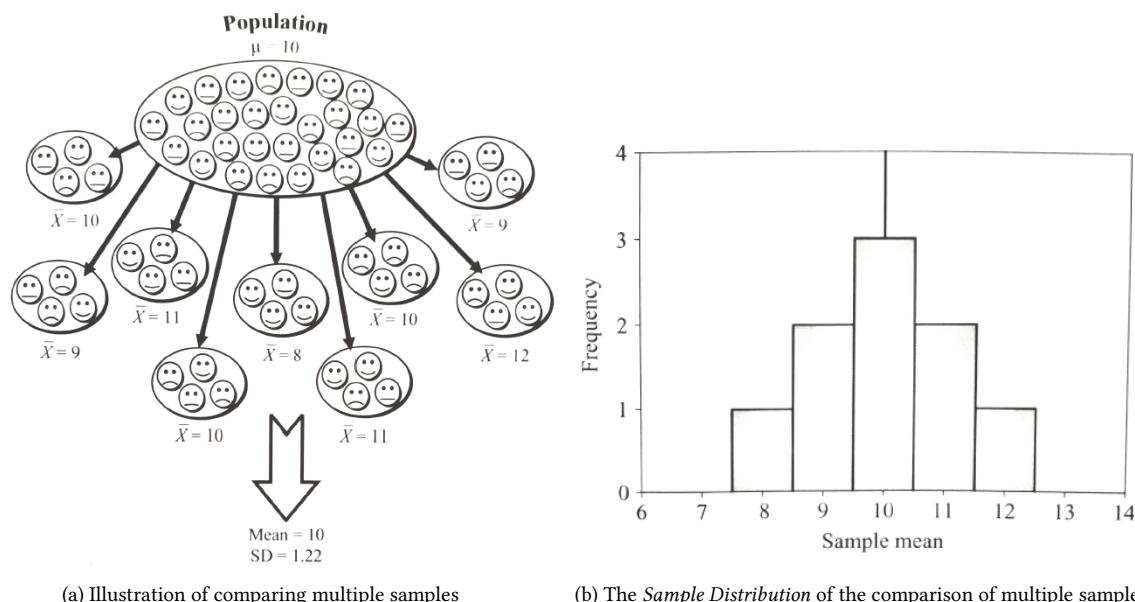


Figure 2.1: Illustration of how to use multiple samples and plot of the sampling distribution.

Comparing Data and Results

The former validation through statistical analysis is only applicable to quantitative data. When working with qualitative data, this is not an option due to the nature of the type of data. It is possible to quantify the qualitative data, but this process often simplifies the data too much.

Simplifying the data results in losing some of the details that qualitative data contains. While checking for distributions can be perceived as a more systematic and rigorous verification of the sample, the opportunity to compare results always exists. Verifying the sample with qualitative results would require consistent results between samples. Further, the individual analysis of the results from a sample should also end up being consistent.

Comparing qualitative data is not easily done, as data entries are very dependent on subjective aspects. Qualitative data is often very detailed, and it is not likely that two identical data points will be present in the raw data taken from two samples, e.g., two identical quotes.

As raw results are likely to differentiate, a stronger alternative would be to analyze the data. Comparing the outcome of the analyses would diminish some of the subjectivity that qualitative data consist of. One way to obtain the analyses is to analyze the results for each sample with Yin's Five-Phased method (Yin, 2015) (for reference, see Appendix B - *Assessment of Methods for Qualitative Analysis* on page 102). The chances for having the same data (i.e., identical quotes and observations) might not be as likely as finding correlations in a quantitative data set. With the analyses, it is preconceived that categories and themes created from the analysis might show up in both data sets from the two samples. The results of the analyses will show if the participants mention the same aspects across samples. Such results allow for a vague verification of sample-matching, but only if one of the samples are from the target population. If the same researcher performs the analysis on the two samples, there is, however, a possibility for biases. The biases concern carry-over effects and the possibility of the researcher either subconsciously or consciously striving to obtain identical categories even though there might not be any.

Disadvantages of Sample Verification

The most eye-catching disadvantage is that these types of verification have to be done in retrospective. Furthermore, both approaches require results from the target population. In the academic world these strategies for verification might be applicable, but in business' it might not be the case. No business is going to conduct a study on two samples with the goal of comparing them to examine how they compliment each other. A business most definitely would not test on the target population to verify results from a sample drawn from another population. Using any of the methods for verification requires that at least one of the samples have been drawn from the target population. If the evaluation already has been conducted with a representative sample, the verification might not matter, as the goal of the evaluation most likely is already fulfilled.

Only having retrospective options for verification, forces researchers to consider the type of sample before evaluation. Without consideration of the sample, there is no guarantee that the sample matches the target population. No considerations could be imagined to result in

high variances. The high variances cause low accuracy of the data which diminishes the result and makes them useless. The low accuracy further disables the opportunity for generalizing the results.

2.4 Differences in Qualitative and Quantitative Approaches

The differences between qualitative and quantitative approaches are often divided into two categories because their aim in research is not identical. The same differentiation in approach is also applicable for sampling in quantitative and qualitative studies. In Appendix C - *Quantitative and Qualitative Approaches* on page 111 a thorough description of methods used for sampling in both quantitative and qualitative approaches is found. Quantitative research aims for testing and answering pre-determined hypotheses and obtaining generalizable results. Qualitative research, on the other hand, has a goal to provide insights and understanding of psycho-social and psychology issues of higher complexity. Therefore qualitative research is better at answering humanistic '*why?*' and '*how?*' questions (Marshall, 1996, p. 522). The differences between the two approaches are more than just general aim. Marshall (1996) listed a table, to give insights into the differences in aim between qualitative and quantitative research. In Table 2.1 the alleged differences are given.

Table 2.1: Marshall (1996) view on the differences between qualitative and quantitative research.

	Quantitative	Qualitative
Philosophical foundation	Deductive, reductionalist	Inductive, holistic
Aim	To test pre-set hypothesis	To explore complex human issues
Study plan	Step-wise, predetermined	Iterative, flexible
Position of researcher	Aims to be detached and objective	Integral part of the research process
Assessing quality of outcomes	Direct tests of validity and reliability using statistics	Indirect quality assurance methods of trustworthiness
Measures of utility of results	Indirect quality assurance methods of trustworthiness	Transferability

Based on the claimed differences between the two areas of research Marshall (1996) emphasizes why *Simple Random Sampling* (SRS) and possibly other variants thereof is inappropriate when sampling for qualitative studies.

In general, Marshall (1996) finds SRS well defined and rigorous, but at the same time finds that this technique is not useful for qualitative studies. He reasons this by the differentiation in the aim of quantitative and qualitative research. The random sample provides the best

opportunity to obtain results generalizable to the target population but that it is not suitable for research aiming for developing a higher understanding of complex issues related to human behavior. Marshall (1996) identifies both theoretical and practical reasons for this. Overall he describes four reasons for not using SRS.

Due to the nature of qualitative research, a small sample of participants tends to enough. With small samples, the chance for sampling errors is likely to be so high that biases are inevitable. A sampling error would typically be, that the participants in the sample are not representative of the rest of the target population. Secondly, Marshall (1996) is under the impression that studying complex human behavior often leads to not knowing the characteristics of the target population. The lack of characteristics suggests that it is possible to draw a random sample, but not possible to ensure that the sample is drawn from the right population - or at least not generalizable.

The third reason for not using SRS in qualitative research regards the distribution of data collected from the sample. While random sampling of a target population probably produces a representative sample, there is in qualitative research no current methods to prove participants of the sample to have identical values, beliefs or attitudes. As the nature of qualitative research relies on these very subjective characteristics a normal distribution is not obtainable, thus making SRS and other probability sampling techniques inappropriate for qualitative studies (Marshall, 1996, p. 523). The final reason is both the researchers and participants skills. By sociologists it is well recognized that people are not equally good at understanding, interpreting and observing their own and others behavior. Also, it is by qualitative researchers also recognized and acknowledged that some participants are '*better participants*' as they provide better insights and somehow articulates in a way that creates a better understanding (Marshall, 1996, p. 523). Finally Marshall (1996) says:

"Choosing someone at random to answer a qualitative question would be analogous to randomly asking a passer-by how to repair a broken down car, rather than asking a garage mechanic - the former might have a good stab, but asking the latter is likely to be more productive"

(Marshall, 1996)

2.5 The Selection of a Sampling Method

As the area of sampling has been described there is still no answer to how to pick a sampling method. Furthermore, it seems that there are many other aspects than just the sample method to consider. The choice of sample is affected by many parameters and obtaining a good and representative sample is not as easy as thought. (Kothari, 2004, p. 56) lists some points to follow when picking a method for sampling.

List of considerations

The below areas should be considered and requires attention. The steps cannot be seen as isolated and do not necessarily have to be followed sequentially since some of the areas will be determining the options for other areas. The list should instead be perceived and used for defining the sample. The definition of sample is not only the selection of an appropriate method but instead a 'design' of the sample. The design consists of all the parameters that potentially influence the criteria of the sample.

Type of Population

No matter quantitative or qualitative approach, one has to define the objects of the study and thereby also the population. The population can be finite or infinite, where finite refers to a specific population defined by certain characteristics and infinite refers to a limitless population from where all and everybody can be used.

Sampling Unit

It has to be decided where a sample is taken. Demo- or geographical factors could determine this, but it could also be a social construct such as family. Typically one or more of such sampling units are selected.

Sampling Frame

The *Sampling Frame* is the definition of users who are going to enter the sample. It is important that the sampling frame is comprehensive, correct, reliable and appropriate, as it is necessary for the frame to be representative of the population. Defining the population entails identifying all characteristic and attributes that could have significant influence in the evaluation.

Size of Sample

While this is not covered in the general description of sampling, this is a crucial decision to make, as this decision constitutes an optimum sample. The size of a sample should not be excessively large, nor too small as it has to fulfill the requirements of being efficient, representative, reliable and flexible. While small samples also can provide good answers, they have a tendency to have less precision than larger samples, which is especially true if the variance

of the target population is high. Larger variances in the population should be considered, as this most likely need to be countered by using an even larger sample.

Parameters of Interest

It is imperative to consider what the actual area of interest is, as this contributes to the decision of whom to use and how many to use in the sample. Area of interest could be estimating the proportion of persons with some characteristics in the population, but could also be a specific sub-group of the target population.

Budgetary Constraints

Until now economical aspects have not been included in the process of picking a sample nor the method of sampling. In the perfect world economic constraints should not exist, but unfortunately, the available budgets for a given study will be a constraint for the possible options, i.e., what kind of users and how many.

Sampling Method

In the end, after all of the above have been considered, one can find a method for reaching a sampling fitting to the criteria and constraints which have just been identified. It is important to consider what method would produce the smallest sampling error and what method that best would help obtain a representative sample.

Typical Causes of Sampling Errors

While the above list of considerations should solve some of the issues regarding sampling, there might still occur incorrect inferences, such as systematic biases and sampling errors for unknown reasons. Sampling errors are random variations in the sample and will typically be solved by increasing the size of the sample - this does, however, increase the cost of the study. Systematic biases are possible to correct if they get detected, but in some cases, the study would have to be re-conducted. Kothari (2004) identifies that a systematic bias usually is caused by one or more of the following parameters:

Vague sampling frame

A vague sampling frame happens, if the people used in the sample was not defined well enough to represent the population.

Defective measures

Could be a bias in the survey or interview, but could also be a physical measuring that is systematically measuring wrong.

The Indeterminacy Principle

When observing people doing a task they might be better or worse on some parameters, due to the individual acting differently when knowing they are being observed. Kothari (2004)

mentions that time for a given task might increase when being observed, but one could argue that the individual also is more thorough, hence fewer mistakes are made.

Natural bias in the reporting of data

When people are included in studies, they sometimes try to figure out what the researchers are looking for, thus making them answer based on what they think researchers wants to hear and not their own opinion. It is also a tendency that people underestimate and overestimate depending on the context of the study, e.g., underestimate their income when related to taxes but overestimating when in context of social status.

2.6 Summary of Sampling

Theory of sampling does not explicitly provide one with knowledge of what to sample in a given study. Instead, the literature seems to provide guidelines and advice for what to consider in the case of sampling, but it also seems very situational. However, a census exists regarding what a good sample is. The sample should be genuinely representative of the target population and allow for small sampling errors. In addition to the sample being representative, it should also be drawn with consideration for systematic biases. Finally, the chosen sample should be economical available, and the results from using the sample should be applicable in general with a certain level of confidence. For quantitative studies, it seems that a good and representative sample can be obtained through the simple random sampling of the target population. For qualitative studies, it is a bit more unclear. Everything should be considered, and there is no 'go-to' method. The best advice for sampling, no matter quantitative or qualitative, is to know the population in scope in detail.

Additionally, literature distinguishes in methods concerning what approach that has been chosen. It is reasoned by quantitative and qualitative having different purposes in research. Thus different kinds of sampling techniques should be used. Even though this differentiation exists, literature seems to interchange terms or have multiple descriptions for identical procedures.

The general theory upon the subject of sampling is quite confusing. Many overlaps, confusions and interchangeable terms make it hard to know why, how and when each method should be used. Furthermore, the theory seems to perceive sampling as one thing: a genuinely representative sample, from a target population. There is no perception of sampling being a sliding scale, where a variety of aspects defines the type of sample. There is only the cemented argument that a sample must be drawn from the target population. While this knowledge surely is useful, it does not provide any insights for answering the research questions. Therefore, other ways of investigating the area of sampling are chosen.

Chapter 3

Empirical Data on Sampling

Theory heavily suggests a sample drawn from a target population, but none of them justifies how one would do if this is not possible. Only the convenient sampling method seems to be applicable in such scenario. The description of the convenient sampling method tells that it is the least rigorous of them all, which does not favor any evaluation. There is a need for a methodological process of obtaining a sample that sufficiently can substitute a sample from a target population. It is envisioned that alternatives exist, as none of the methods applies to the scenario of an unreachable population.

Through investigation of previous studies and interviews with experts working with usability and user experience, hopefully, some information on how to approach sampling will be found. Investigating previous studies should help to identify when and how to use different kinds of sampling methods. Talking with people outside the academia should give insights to how sampling is done in business-related situations. Investigating sampling in a business' might give a chance to triangulate the information obtained through previous studies. A triangulation and comparison of the information and results obtained through this chapter lead to a discussion. Hopefully, this discussion can provide some answers to how to approach sampling. Therefore, both exploratory, normative and summative studies are in scope for investigation. In addition to '*when*' and '*how*' it also desired to investigate the relationship between results and type of sample used.

3.1 The Value of Context

In usability studies, context is said to have very high importance. The importance is normally reasoned by contextual usability problems only being identifiable in evaluations taking care of context. Additionally, Marshall (1996) states that the context of a study has high importance, due to temporal, spatial and situational influences.

Even though it seems that context has a profound impact on what usability problems become identifiable, there are areas where it is wise to do laboratory evaluations. Kjeldskov et al. (2004) investigated the value of field evaluation compared to laboratory evaluation. The incentive for doing so was that literature showed most studies for mobile HCI research projects were conducted in laboratory settings though researchers argue in favor of the field-based evaluations due to mobile-systems being context-dependent. They addressed the issue, as the field-based evaluations are difficult to conduct and resource intensive. Finally, the added value compared to a laboratory setting is unknown.

The results from the study are shown in Table 3.1 and Figure 3.2. They found significant differences in the identification of cosmetic and serious usability problems, which favored the laboratory setup. The identified critical usability problem was however not significantly different between the two methods. Depending on the goal of the evaluation, it could be that laboratory evaluations can be favored, as they are less extensive to plan and conduct. Kjeldskov et al. (2004) does, however, mention one specific problem that was only found in the field evaluation, thus showing that laboratory evaluations might miss certain aspects that only real-life context will be able to show.

	Laboratory (N=6)	Field (N=6)
Critical (N=8)	8	7
Serious (N=19)	18	10
Cosmetic (N=10)	10	6
Total (N=37)	36	23

Table 3.1: Distribution of identified usability problems for each type of evaluation

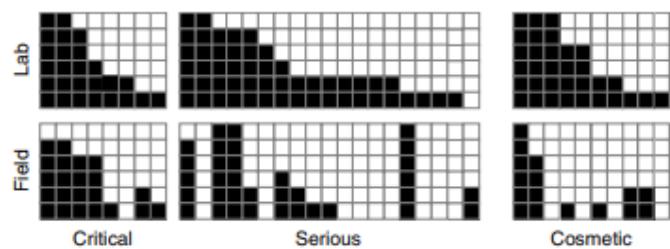


Table 3.2: Distribution of identified usability problems visualized. A column represents a usability problem related to the number of participants identifying the particular problem. A black box indicates the identification of a usability problem.

Kjeldskov et al. (2004) condensed the results it into four key findings. They do however disclaim that their study suffers from certain limitations. The prototype they evaluated and the health care context it was put in was potentially influencing the results. Additionally, they considered that other characteristics would influence different domains; thus coherence of usage of the prototype and context might be weaker or stronger. However, the four key findings from the study were:

1. Little added value of taking the evaluation into a field condition.
2. Lack of control undermined the extendibility of the field condition.
3. Both the lab and the field revealed context-aware related problems.
4. The clip-on camera facilitated high-quality data collection for mobile use.

Interpretation

The findings suggest that the context of some evaluations, might not be as important as anticipated and as literature describes. As many parameters can define context, it is interpreted that when asking users to do different assignments, typical usability problem will be identified, unless the problems are strictly affiliated with a specific context. A potential solution to avoid this would be to analyze tasks the participants are asked to do. Taking this approach would identify how context-aware the given assignment is. It might also open the possibilities of relaxing on screening criteria for participants, where the important characteristics of such population are very little context-dependent.

Contextual evaluations, such as the field-evaluation also introduces areas not controllable by the researchers. The number of variables also has to be considered up against the goals of the evaluation and the intended data collection. A controlled environment for testing often enables a stable framework for investigating correlations between variables.

3.2 The Influence of Location, Participants, and Resources

Context is one parameter that can influence the results from a usability test. Kjeldskov et al. (2005) further investigated the different methods for usability testing in another study. The goal was to clarify the value of usability methods through an extensive multi-method evaluation of a mobile guide. For the issues regarding sampling, evaluation of mobile guides are also vaguely described by literature. Therefore, identical issues are encountered within this area; how should one approach the problem, what are the methodological challenges and what are the pros and cons of different usability evaluation methods? (Kjeldskov et al., 2005, p. 52). This lead Kjeldskov et al. (2005) to investigate the following three aspects of usability testing:

1. In-situ or in-vitro

In-situ or in-vitro concerned the choice of field evaluation or laboratory evaluation. The main issues were by Kjeldskov et al. (2005) defined as field evaluations providing greater context but more troublesome to conduct. Laboratory evaluations could provide higher control and quality of data, but usability laboratory setups could be inadequate to simulate context of use.

2. Users, surrogates or experts

Users, surrogates or experts concerned the choice of participants. One concern is that both field and laboratory evaluations possible consist of a task that can be de-contextualized. Furthermore, both evaluation methods focus on evaluating users' interaction, which is time-consuming and challenged by participants being unfamiliar with the process. Kjeldskov et al. (2005) refers to solutions involving expert-based evaluations such as *heuristic evaluation* and *cognitive walk-through* (Nielsen and Molich, 1990; Wharton, 1994). The provided solutions are deemed beneficial due to giving evaluators guidance on the subject. By literature, a critique is given for these approaches as they identify fewer usability problems compared to methods involving users.

3. Exhaustive or discount analysis

Exhaustive or discount analysis is considered due to the intensive amount of resources required for the analysis of usability evaluations. Kjeldskov et al. (2005) questions the relative value of exhaustive analysis, as the value of doing so is largely speculative, but still considered the gold standard for analyzing usability evaluations.

To investigate the above Kjeldskov et al. (2005) used *grounded analysis* (Strauss and Corbin, 1997). Also, to enable a comparison of each method, a compilation of the results was performed. The compilation followed a predefined method, to ensure that the process of compiling and comparing was rigorous and that the results were credible, dependable and conformable (Lincoln and Guba, 1986). The results from the study where obtained through analyzing and comparing these compilations across the methods of usability testing. Table 3.3 and Figure 3.1 shows the results.

	Field Evaluation	Laboratory Evaluation	Heuristic Walk-through	Rapid Reflection	Total
Critical	4	4	4	4	5
Serious	7	6	6	5	11
Cosmetic	2	3	3	4	6
Total	13	13	13	13	22

Table 3.3: Distribution of identified usability problems for each method



Figure 3.1: Distribution of identified usability problems. The figure shows that no single method were able to identify all usability problems.

From the table, it is possible to deduce that all methods identify the same amount of usability problems in total, but that these are scattered in between the three categories. It is also deducted, from the figure, that some methods identify problems that the other methods do not. Finally, all methods identify more than half of the usability problems identified from the four methods (13 identified problems out of the total 22 identified across the four methods).

For the general view of the table and figure, the four methods do not variate much when looking at identifying critical problems. A serious problem, however, shows that the heuristic walk-through and rapid reflection brings variance into the identified usability problems. Moreover, the one bias of using heuristic walk-through was observably not present in the evaluation, as the amount of '*cosmetic noise*' was not inarticulately high which is the main critique of the method. Kjeldskov et al. (2005) states that the tailored heuristics used in the study might have diminished this type of bias. Kjeldskov et al. (2005) discusses the importance of the considerations, they did before choosing methods.

When revisiting the concern of in-situ or in-vitro, this study (Kjeldskov et al., 2005) states, that the use of in-situ was their insurance against ignorance in the absence of an understanding of what the context of use was or might become. It also states the in-situ approach (field evaluation) have provided insights that the team might have been ignorant of or might have passed unremarked. Based on the findings Kjeldskov et al. (2005) states that field evaluation is not to be replaced by other methods, but instead should be implemented as a broader collection of data unless there are practical or logistic challenges that make it impossible to access the product in-situ.

On the subject of *users, surrogates or experts* there is not much to withdraw. Instead Kjeldskov et al. (2005) refers to a more general discourse of the discussion of the issue expert versus user-based evaluations. During the conduction of the study, the facilitator observed some evidence of experts being able to overcome the credibility of low fidelity prototypes were present.

The concern regarding *exhaustive or discount analysis* Kjeldskov et al. (2005) states that what matters is the objective of the study. The discount approaches should be used in early

phases where user input applies to product development. The application of the exhaustive approach is deemed to be when the scope of research is beyond design input. The argumentation for this was that the nature of the exhaustive approach also focuses on safety-critical and business-critical aspects (Kjeldskov et al., 2005, p. 63)

Interpretation

Once again research shows how context is not everything when conducting usability tests. It is however evident that methods involving end-users (field- and laboratory evaluation) might help identify different problems than methods involving experts (heuristic walk-through and rapid reflection). The use of experts seems to cover some of the same problems as the end-users experience, where it could be imagined as a temporary solution if end-users are not available. No matter method, a bare minimum of end-users seems to be preferable, to make sure specific usability problems are not coherent with the specific type of users - and thereby end-users functions as insurance against ignorance.

3.3 Alternatives to User Inclusion

When talking about usability testing, some previous studies have also troubled to take a sample from the target population. Many of those, try to rely on the sampling methods associated with these *hard-to-reach* populations. Additional requirements also appear in some of these studies, as some of the hard-to-reach population include more ethical aspects (Kennan et al., 2012).

Shaghaghi et al. (2011) did a review of the literature concerned by these methods, and the conclusion points out, why the methods themselves are not sufficient. It does not take the advantages or disadvantages of the methods into account, but instead Shaghaghi et al. (2011) states that the successful use of any of the methods designed for hard-to-reach populations is solely dependent on knowledge about the population. The knowledge needed for using any of the methods consist of the characteristics of the population and the sub-groups within. This need for knowledge is persistent through most, if not all, studies concerned with sampling. There do however exist alternatives to the user inclusion. It does not replace the need for actual usability testing with end-users, but it helps the development of the product. The alternatives are personas, identity model, user segments and user profile. All these alternatives have one common goal; they are to represent the characteristics of the end-user. They reflect a population's needs and goals, which possibly can be incorporated in the phases of development where it might not be possible to recruit a sample from the target population (Cooper et al., 2007; Dayton, 2003). The existence of these methods and the issues with the current sampling methods for hard-to-reach populations can suggest that it is possible to work around the need for the target population. It is however not evident how this should be approached nor if it is possible.

3.4 Interviews with B&O and Technolution

Revising studies have proven to establish a small opportunity to relax on criteria of samples. It is however very dependent on the domain of research and type of method but suggests that alternatives do exist. To establish further understanding on the subject of sampling interviews with two people from the industry is to be conducted.

Method

As the objective of the interviews is to obtain empirical data on the issues within sampling, an open-ended, unstructured interview is chosen. Further reasoning for the choice of method is that there is no particular expectation to the format of answers nor a wish for obtaining data on particular questions. Instead, the only desire is to obtain knowledge on the subject of sampling. The method allows the participant to propose and identify solutions or issues that have not been considered by the author. Other methods within the genre of interviews would possibly limit the participant; thus the open-ended, unstructured interview is chosen to enable the participant best to share knowledge (Rogers et al., 2011, pp. 228-292).

The desired type of empirical data to be obtained through the interviews is mainly exploratory. From the data, it should be possible to map how the issues of sampling in hard-to-reach (and impossible) situations are present in different corporations and how the issues of sampling are worked around. Besides learning about hard-to-reach samples, the goal of the interviews is to learn about how sampling usually is practiced in corporate settings. The desirable results are hopefully to be achieved through discussion of the subject of sampling. While the interview is decided to be open-ended and unstructured, it ensured that both interviews address the domain of sampling and the issues relevant to the study. Ensuring some structure is done by a focus question, which is; '*How does one approach sampling when challenged by confidentiality and hard-to-reach population issues?*' The questions will not be asked word-by-word but will function as an overall theme for the interviews. It is to be the central point of discussion within the domain of sampling. Acknowledged, the answer to the focus question covers a variety of aspects which hopefully will be discussed in-depth through the interviews.

Type of analysis

The typical outcome from an open-ended, unstructured interview consists of primarily qualitative data with the addition of a few quantitative metrics, i.e., the age of the participant. The type of interview chosen collects the raw qualitative data through audio recordings and notes taking during the interview. Video recording can also be used in situations where needed. The initial step of data processing is to transcribe the data deduced from the interview. The transcript covers the whole conversation from start to finish. The further processing of

data consists of several options, and the selection of the method for analysis is very dependent on the goal of the analysis (Rogers et al., 2011, pp. 270-292). For choosing the most fulfilling method, some methods have been described and assessed with considerations to resources and potential outcome and goal. This can be seen in Appendix B - *Assessment of Methods for Qualitative Analysis* on page 102.

Due to the goal of the interview, it has subjectively been chosen to apply a Meaning Condensation Analysis, based on the assessment of the different methods. In the appendix referenced above, a further description of this type of analysis can be found. The essentials of the analysis are to reduce complexity, thus being able to have more concise and meaningful statements and themes. Both interviews have been transcribed before any analysis is applied. The focus of the transcriptions has been to obtain a text understandable for the reader. Thereby, the transcription is true against what the interviewee has articulated, but aural information and nonsense have been omitted, to increase the readability. Additionally, the transcript omits some of the start and end of both interviews, as this information is not about the subject of sampling. A few indications in the transcripts were in some cases judged necessary. Overlaps and interruptions are indicated with 'l'. Empty parentheses ')' indicates an inaudible word, and a word in parentheses indicates a possibility of what the word could be (Silverman, 2015, p. 449).

3.5 Senior UX & Usability Lead, Tina Øvad

Tina Øvad is an expert on the subject of user experience and usability and holds a Ph.D. in Engineering Psychology from Aalborg University. At the moment Tina is employed at B&O where she daily works with user experience and usability. Tina was interviewed to investigate how problems within sampling are experienced outside of the academic framework. The interview was conducted on Skype, lasted 1 hour and 37 seconds and, mainly addressed the subject of sampling. The transcription of the interview is found in Appendix D on page 119.

Through the meaning condensation analysis, a total of 41 themes (*also referred to as unit*) have been identified. In the following descriptive statement, a parenthesis with U# indicates the actual theme from where this is interpreted. Some of the themes overlap, due to subjects being revisited throughout the interview, whenever a new perspective could be applied. In short, Tina related to the issues within sampling and gave additional input to how the subject could be approached. Furthermore, several important aspects related to sampling were identified, hence making the whole issue of recruiting and sampling even more complex. Tina, in general, stated that context is an important aspect and that many aspects have to be taking into consideration when finding a sample. The below statement is derived from the analysis of the interview in table E.1 in Appendix E on page 140.

Statement from Analysis

At the beginning of the interview, it was stated that sampling could be approached differently in some cases (U2, U5). In handling tasks, it might not be necessary to have actual end-users as some human-factor evaluations only require general handling and do not require the sample to have specific prerequisites. Not needing actual end-users is quite a statement and is only valid as long as the handling required is not influenced by specific knowledge limited to the target population or if the target population has unique prerequisites or disabilities (U3). From a practical point of view, some evaluations concerned with handling will be able to diminish the requirements for a sample drawn from the target population. It is however not that easy to conduct user experiences research with a substitute sample, as this kind of evaluation targets more cognitive influenced aspects (U2, U5, U7, U20, U21).

In addition to this preliminary prediction, the goal of each evaluation could undergo analysis, as different goals of evaluation set different criteria for the sample. Through analysis, it may be possible to widen the criteria for a sample. Additionally, such analysis could provide a weighting of the influence of each criterion, making it possible to select the most influencing factors (U4, U7, U17). This analysis could also provide one with enough information to identify a substitute population which fulfills some of the influencing criteria (U8, U17).

While some criteria might be easy to obtain through a substitute sample, others might be harder. It was suggested to avoid or minimize the impact of these criteria on the evaluation to allow for valid results and fewer difficulties obtaining a substitute (U9, U10). To achieve this, one has to know the characteristics of the target population but also has to keep in mind what limits and constraints there is by using a substitute sample.

As the substitute sample is an approximation of the target population, it is not sure that all criteria can be fulfilled in a single evaluation. To accommodate that all sampling criteria are present, an evaluation could be performed multiple times with participants who singularly does not cover all criteria, but together does. Furthermore, an evaluation could be split into smaller fragments, e.g., the handling part is done with one sample and another task where other criteria are essential is done with a second sample. A number of these fragment-evaluations would then combined represent a full study with all sampling aspects that have an influence. (U12, U13).

Evaluating this way requires awareness of context and the objectives of the evaluation. There is a need for awareness in all phases of evaluating with a substitute sample. In general, more effort has to be put into the evaluation. Both in the phase of designing the evaluation and picking the sample, but also being aware of the impact the evaluation design and type of sample has in the analysis of the data collected throughout the evaluation (U14, U15). Another impact that potentially could be neglected is the influence of information given to the substi-

tute sample. Feeding context into a task or situation has to be considered, as some information might be able to skew the data. An example could be telling a person with diabetes to inject himself five times a day - no issues, he accustomed to that. Telling a naive user from a substitute sample, who is not accustomed to that could lead to skewed data, as this participants reaction to the information is predicted to be stronger (U16).

Related to the analysis of the target population, personas or other user-centered methods could be created (or even a 'persona' for the disease). It could provide knowledge on what kind of information to give, but it would most definitely also identify the situations where a substitute sample never will be appropriate to use (U18, U20, U21). Besides identifying if a substitute is suitable, it would allow gaining insights into the users' needs, motivations and attitudes. The insights could then further be used in development phases, hence reducing the need for including the user in some situations. (U19).

Due to the various aspects related to the choice of sample, a natural subject was to discuss solutions. At first, a two-dimensional matrix was suggested, where the first dimension would be the method of evaluation and the second would be the type of sample allowed (U6). It was however later identified that such a tool has to be very general (U22), as the context of the study decides sampling options. The various context aspects lead to the idea of increasing the dimensions of the matrix. It should, however, be focused on the sampling aspect and not the method, i.e., it should not dictate that qualitative methods were the best, when having a low number of participants, but instead suggest alternatives to sample, almost no matter method (U23, U27).

Another subject discussed in the interview was keeping products unexposed for users. B&O used to evaluate products in-house and rely on expert knowledge. Tina does, however, mention that this they are getting rid of this approach. The argument for getting rid of this outdated approach is that the value of user input is perceived higher than gut-feelings on how the product should be (U24). As they are moving towards a new way of handling their evaluations, Tina states that professionalism has to be present when sorting and analyzing the input. In short '*the more eyes, the merrier*' in all phases of development leads to informed decisions and thereby an informed design (U28).

Users should especially be included in the early phases where user needs are established and worked into the product. Features and design get more attention in the later phases compared to the early phases. Users could also be included in later phases of development to ensure that the design is on track and that the product fulfills the user needs (U25). At B&O this has not been the case previously. B&O has previously done some user evaluations earlier and pinpointed problematic element in the design. Unfortunately, these findings have not been reacted upon or taken seriously. This was due to ignorance of the results obtained through evaluation (U31).

Including users in early phases can, however, be problematic, due to the fidelity of the prototypes. Low fidelity prototypes can make it difficult for the participant to relate to an actual product. Hence an evaluation without considerations to product fidelity could provide questionable input. Additionally, the low fidelity can exclude some aspects, due to the missing functionality (U26). As for the change in the approach it is wanted to listen to the users and have a 'People before Technology'-approach (U29). This approach eliminates the development of products forcing users to adapt, and instead forces the development of the product to accommodate user needs. When using a specific sample for an evaluation, it has to be considered that the segmentation of the users can change depending on perspective, i.e., marketing or UX perspectives. One might argue that an archetypical segment could be a solution, but this could result in a segment too broad. Using the archetypical segment for drawing a sample would then likely result in a sample not representative of the target population (U30).

In general, it seems that UX and US are perceived more serious. The maturity of UX and US have increased over the years and is now perceived as an effective evaluation by business' compared to the past. It was through the interviews found that previously UX and US were perceived unnecessary until the point where a product failed miserably due to the lack of evaluations and user input (U32). While user experience and usability evaluation are not viewed that way anymore, there are still issues. The process of recruitment is still perceived as a constraint in many cases, but it is hard to identify why. A general perception is that most people are more than willing to participate when asked (U34, U35). As willingness is not an issue, Tina pointed out that the character of the issue with recruiting might be more practical and potential solution might be to increase the accessibility of users. Furthermore, it could be due to concerns for others, e.g., concern for wasting people's time (U36). One way to avoid the issues of recruitment is to outsource it. Outsourcing the recruitment requires the use of third-party agencies which generally slows the process of recruitment. Additionally, it increases the cost (U33)

In addition to all of the above, researchers within the field of UX and US might be *prima donnas*. We have been taught to do exactly as the method states, and the possibility of relaxing on some criteria often does not exist. This lead to a way of thinking, stating "If criteria are not met, it is not worth doing" (U37). The academic methods for conducting user experience research and usability studies are not conforming to the needs of a business. A possible relaxation on criteria might not lead to results categorized as academically valid, but still valid enough to be the basis of an informed decision (U38). With this in mind, it seems that a shift in paradigm is happening. The perception of user studies is slowly changing. The change in paradigm is that the categorization of user research is a sliding scale and not a cemented method of doing research (U39).

As for this '*shift*' in paradigm, it can be hard for any researcher to go up against 'the norm' with a non-academic approach and vouch for valid results, without actual proof.

3.6 Senior Design Engineer & Human Factors Specialist, Morten Purup

Morten Purup also participated in an interview to clarify how sampling is approached where he works, and how he generally relates to the subject. The interview was conducted through the phone, as Morten suggested this, for the interview to fit his calendar. The interview took 44 minutes and 50 seconds. Identical with the interview conducted with Tina the subjects addressed mainly referred to sampling. The transcript of the interview is in Appendix D on page 132.

Through the meaning condensation of the interview with Morten, a total of 22 themes were created. As with Tina, Morten emphasized that context and goal of evaluation will be deciding the possibilities of using a substitute sample. Additionally, the task analysis is given as a possible alternative solution to, how situations where using the target sample is near impossible. A prediction from Morten is, however, that the target population will be a fundamental need in some phases of development. The statement below is constructed from the analysis of the interview in table E.2 in Appendix E on page 152. As with the previous interview, all themes are referenced by *U#*. Please be aware that the units are taken from Table E.2, hence not referring to the same themes as the units used in the previous interview with Tina.

Statement from Analysis

It became evident that the use of a wrong sample produces results not valid enough to ensure a safe and effective product conditionally. Morten typically delivers validating results to the *Food and Drug Administration* (FDA) where representative users are a must. The burden of proof to the FDA contributes to the fact that no questions are asked regarding the expenses related to the process of recruitment (U2). Furthermore, the burden of proof for the FDA is high, thus creating demands for the type of evaluation and sample (U12, U15, U16). While the burden of proof in some cases sets a demand for sampling from the target population, small changes can typically be tested on substitute samples. To prove that a substitute sample is adequate to a sample from the target population, some argumentation has to be applied. The argumentation has to prove that the design- or concept changes are not affecting the use of the product. Furthermore, it should be proved that the intended evaluation does not demand specific prerequisites only found in the target population (U16, U18). Generally, evaluations with the goal of validation should include end-users while '*less important*' decisions can benefit from a substitute sample, e.g., internal tests (U16).

However, situations do exist where substitute samples can be used. Evaluations can benefit from the use of a substitute sample if it is proved that the substitute sample has the same prerequisites. It also has to be proved that the evaluation can be conducted under the same

conditions, e.g., a field-evaluation where context also is essential. These situations are also based on the type of evaluation, hence making it a goal-based sampling for evaluations (U3, U5). A pitfall in the use of substitute samples is when '*know-how*' in a given situation is required, that a non-representative participant will not be able to know (U6). While one could argue that this information could be given otherwise, it was discussed whether it would be realistic enough to provide this know-how and still obtain valid results. This calls for consideration on where and when knowledge should be passed down, to enable the use of substitute samples (U7).

Even when a substitute sample is available and considered a genuine alternative, one might consider not doing so. End-users will often, if not always, be able to provide richer data useful for the researchers. This richer data can be anything, but in many cases, it will be data not accessible through a substitute sample (U21, U22). A substitute sample will however always be preferred if it is a genuine alternative and the target population is inaccessible.

As with the previous interview, it was emphasized that the need for sampling from the target population could occur in any phase during development, as the goal of the evaluation dictates the sample (U4). Another approach to take, to avoid using expensive resources on recruiting from a target population, is to take an analytic approach. During such approach, all tasks are broken down into discrete steps and analyzed by an expert. This approach should, however, be paired with usability tests and other methods for empirical data collections, as these compensate and not replaces each other (U8).

One reason for pairing and triangulating data from the different methods is that the analytic approach cannot identify unpredictable situations - which will happen during any evaluation with users (U9). Adding a layer of analysis is possible to the task analysis, which is the use of mental models; which mental models does the product afford and how would these be affected by, i.e., different personalities and ideologies (U10). The added layer of analysis could easily be a combination of the previous personas mentioned by Tina and the task analysis itself.

The problem of reaching hard-to-reach populations was also present in other places. The conditions for a study is not the challenging part, but the recruitment of representative users is. It is due to the limited number of people with a specific condition (U11). One way of accommodating the challenge is naturally to increase the budget for recruitment (U13). When traveling between participants are necessary a crucial factor is to control all variables. These need to be identical for each participant, which also implies identical facility conditions. It is the researcher's job to ensure that all these needs and demands are ensured (U14). While these challenges of recruitment exist, there are situations where it is impossible to use a substitute sample.

This is due to experience and knowledge that cannot be given otherwise (U19). Manipulating feelings, opinions, and knowledge are near impossible in all situations. There do, however, exist situations where the conditions can be manipulated, e.g., stress. One might be able to identify areas where manipulation of some sort can be applied, to simulate a given aspect (U20).

On the subject of keeping a product in secrecy, this can be perceived as a self-inflicted issue. Trust should be put into the *non-disclosure agreement* (NDA), as this should keep the exposure of the design or concept under control. As with the previous interview, a general view is the more feedback, the merrier (U17).

3.7 Discussion of Literature, Theories, and Empiri

The process of investigating sampling in both a theoretical and practical setting has proved differences between the two. Through the interview with Tina and Morten additional aspects of sampling have been identified.

A Flaw in Literature and Theory

In theory, one flaw is the presumptions about random sampling. The presumption is, that if the nature of the population is defined correctly a random sample can be drawn and be representative. Everyone in the population should still have equal chances of being picked. This assumption seems only be applicable in theory. The literature states that a sample can only be representative if the sample also shows the diversity that exists in the population. If a random sample worked in practice, a random sample would also have to take this aspect into account. The random sample does not give any insurance of being representative of all cultures if it really is randomly drawn. It leads to the idea that there will never be a *true* random sample representative of the population, but merely a random sample consisting of the target population. Another flaw is that all mathematical and statistical viewpoints on sampling are based on the presumption of a genuinely random sample. Statistics does, however, seem to be fulfilling with this approach, but it is up for discussion whether this is due to a rightful presumption on random sampling, or it is the benefit of the size of the sample. The chances for obtaining normally distributed data, thus suitable for statistical treatment are often benefiting from larger samples.

Aware of Context

In addition to the presumptions, the literature does not seem to include aspects regarding context nor fidelity of a prototype explicitly. Dumas and Redish (1999) and Kothari (2004) does implicitly imply that context is essential for the evaluation and sample, e.g., participants doing real tasks and that the final results shall be applicable in general. Contradictory, both Tina and Morten explicitly implied that context is essential for the choice of sample. Additionally, they were both concerned with fidelity and goal of evaluation. There seems to be a contradiction between theory and practice. The contradictions are also valid for methods used in the evaluations as well. There is a strict academic process, which in practice is not useful, due to the method not being agile and flexible. The lack of flexibility seems to apply for sampling too, as recruitment in practice is expensive, time-consuming and in many cases troublesome. Thus sampling not being agile or flexible too. The methods for reaching hard-to-reach populations also rely on some luck, as, e.g., snowball sampling relies on previous participants social networks in the target population.

Kjeldskov et al. (2004) further showed how context might not have had that big impact usability testing a product. Besides, they also showed that experts also can identify some of the critical and serious usability problems, that end-users can find. The findings Kjeldskov et al. (2004) obtained are also in coherence with what Morten said about task analysis, which suggests that some steps of usability evaluations do not necessarily have to be conducted with end-users nor a substitute thereof.

In the situations where a user inclusion is necessary the actual goal is to obtain a more convenient sample, still being representative. Morten and Tina both provided suggestions for how this could be obtained. Condensing the provided solutions and comparing it to literature suggest that all aspects of the target population have to be identified and mapped. Other aspects, such as fidelity and goal of evaluation should still be considered at the same time. The goal would be to identify all perceptual and cognitive skills from the target population. Further analyzing the evaluation and comparing what prerequisites the tasks require enables one to compare these prerequisites to the skills of the population. The comparison could potentially allow for an expansion of the sampling criteria if the prerequisites required for the task is not dependable of any of the characteristics and skills the population has. If only some of the tasks require the use of the target populations, a solution could be to change the nature of the tasks, to avoid this demand of the target population. Avoiding the task-induced demands is only possible if the changes introduced to the evaluation does not obstruct the collection of data needed. If the results from the evaluation are not sufficient, the aspects left out will most likely be due to the missing aspects unique for the target population.

Sampling from Target Population, as the only solution

Even though a substitute sample or even experts can replace the need for a sample of end-users, it is important to remember that this replacement is situational. There do exist situations where the only option for a sample is from the target population. Morten emphasized that validation or evaluations of *important* things, is to be done with a sample from the target population. Morten has the same strict burden of proof as Novo Nordisk, as he has to secure both the safety and effectiveness of a product to FDA. As FDA is a prominent authority with high standards, there are limits to what can be done, especially in summative usability evaluations. While the burden of proof potentially can exclude the use of substitute samples, the content and goal of the evaluation also have to be analyzed. If the evaluation requires specific know-how strongly related to experience or feelings, it will be tough to transfer these aspects to a substitute sample, without introducing some bias. The biases could be anything, but when giving information, knowledge or setting the scene, participants will most likely react differently than the target population, if the information is transboundary compared to their own lives.

General Differences of Theory and Practice

Theory and sampling are in practice perceived as two different things, but they do however compliment each other in some areas. All sampling is, and will always contain convenient elements. It might be intentionally or unintentionally, but the literature does not coincide how sampling takes place in practice.

The theory is right that the best sample is one from the target population, but it does not provide any information on what real alternatives that can be used when this is impossible. Arguably, the convenient sampling method is the answer to everything. When none of the more rigorous methods are possible, the convenience method is the last alternative. It does not exactly seem like a satisfying alternative, when accounting for, e.g., the burden of proof. This statement seems to be validated by the argumentation from Morten, where he explicitly explains how an uncontrolled (e.g., too vague criteria for sampling) process can undermine the results gathered, as they do not represent the actual users. Morten luckily had the resources and budget to reach his target population, but there are no alternatives if resources are profoundly limited.

It is unreasonable to think that all businesses have access to their target population. All businesses without access to sample from a target population, would in theory never obtain valid input through their evaluations. In practice, this is not valid and is of course, not always the case. The reason for this argument not being valid is that context and goal of evaluations take part, as the need for sampling from target population seems to be situational and based on these factors

With this discussion of literature and the empirical data collected, it seems possible to provide some answer to the research questions initially stated. Although there still are areas to investigate regarding sampling, it is deemed that further investigation, only will make further aspects surface.

Chapter 4

How to Approach Sampling

Through the revisit of literature and previous studies as well as the interviews, it is possible to give some answers on how to approach sampling, when the target population is out of reach. Answering the interim questions will not validate the use of an alternative approach to sampling, e.g., finding a substitute. Instead, it will hopefully result in ideas for a method that will enable the identification of substitute sampling.

4.1 Answers to Interim Research Questions

The interim research questions addressed one at the time. Each question will be answered separately and lastly summarized together. It is essential to keep in mind that the empirical data collected through interviews and previous studies are very much dependent on the experts own opinions and ideologies. If other persons had been interviewed or other studies had been reviewed, other results might have been obtained. It was through the interviews it became apparent that subjectivity has a substantial influence on the subject of sampling. While the viewpoints of sampling come down to personal beliefs, ideologies and opinions the interviews showed how the burden of proof, collaboration with authorities and the general context of everyday work influences researchers perspective on sampling.

How is sampling approach, when the confidentiality of a product exceeds what is reasonable to expose outside a business framework?

Even though there surely exists situations where confidentiality will not allow for exposure of the product outside a business framework, it will in many cases not be sensible not to. The value of exposing a new concept or design exceeds the value of keeping it inside the business framework. While experts and product owners can give very decent advice and direction for the development, user evaluations will often provide richer data and results.

The exposure of new concepts or designs is more vibrant because user evaluations at the least will provide the same or equal results to what could be obtained through experts and analysis. The chances are that involving users will provide insights not possible to obtain otherwise, hence providing additional data. If it is vital to keep the design in secrecy, it could be argued that features and functionality could be tested in another design. Without exposing the actual design, it would be possible to get an evaluation of functionality and features of the

product. Such approach would allow for testing specific functionality, but it would not be able to test a full concept. A full concept would require everything to be, as it is supposed to. Some concepts might rely heavily on new technology and features. Thereby, this suggestion will not be applicable.

From the interviews is understood that the more input and feedback, the merrier. In this context, Morten also articulated that keeping a product in secrecy and afterward asking for solutions to evaluate is non-sense. It is a self-created issue that one has to accept when deciding to keep a product inside the business framework. Even though there are issues in keeping the product away from the user, there are still options left for identifying usability problems. These involve expert evaluation such as the task analysis or the heuristic evaluation. When using these approaches to evaluate a product, the limits of the method has to be accepted, as they most likely will not produce as valuable findings as user evaluations.

Effect of qualitative and quantitative approaches to sampling

Both the quantitative and qualitative sampling methods share many aspects. No matter what, a sample will always include some aspects of convenience and purpose. Convenience naturally happens, due to necessary compromises, such as location. The convenience aspect is determined by the resources allocated to the single evaluation, where resources can be anything from budgetary constraints to amount of people able to conduct the evaluation. Smaller resource budgets will most likely rely more on convenience due to inaccessibility to the required population. The purposeful sampling will also be present in both the quantitative and qualitative approach, as no one is going to evaluate without a purpose. If the purpose is to evaluate a given product, all researchers should consider the variables and their influences. Influences and variables should be accommodated for in some way if their impact is unwanted in the evaluation. There is no real answer to how to accommodate a sampling issue, without just increasing the budget for recruitment or relying on a total convenience sample. There is a justification for proceeding to develop a method, as sampling is much more than representativeness and convenience.

Even though it sounds like there are no differences between qualitative and quantitative sampling approaches, there is. When conducting quantitative studies, ensuring that the data collected is suitable for further statistical treatment is a must. Most statistics demands normally distributed data or homogeneous variances and some actions should be taken to ensure that the right data is collected. The definition of 'right' data is based on the envisioned analysis. It could be argued that the same applies for qualitative studies, but this is only due to the wish of reaching a saturated set of data., e.g., interviewing until no new subjects are brought up by participants.

As a rule of thumb, qualitative studies need a minimum of 6 people to cover most aspects of a question. In comparison, when a quantitative study aims for normally distributed data, the rule of thumb is minimum 30 participants. For both approaches, the sampling size must be sufficient to ensure conclusions inferred for the whole sample, given the sample results.

When it is not possible to draw a representative sample from the target population

Literature does not have any answers for this. Therefore, the reviews of previous studies were conducted. At the same time, the interviews were conducted. It is fair to say, that if resources are sufficient to reach the target population this should be done and it would be unreasonable not to. In situations where the target population is not available additional resources will not provide any advantages. The demand for a sample from the target population is also based on the goal of evaluation and the method chosen for obtaining these goals often dictates the type of results. Previous research showed that context in some situations would not have an impact. Furthermore, it was shown that conducting contextual, i.e., field-studies there is a chance to increase the complexity of conducting the study. Laboratory studies were deemed adequate in some situations which suggest that knowledge of evaluation possible can help identify when and how a substitute sample should be used.

To obtain knowledge of the evaluation in scope, many aspects have to be accounted for. The needed knowledge can possibly be obtained through analysis. An analysis would enable the identification of tasks where a definite need of the target population is present. Knowing when and where the need for the target population is present implicitly also states when and where the target population is not needed. These are the encounters, where a substitute sample can be used instead.

Knowing where a substitute can be adequate leads to how this should be done. Through analysis of the evaluation, the prerequisites for completing the content of the evaluation should be present. It is as a result of this possible to compare the characteristics of the target population and the prerequisites of the evaluation. If the prerequisites do not show in the characteristics of the target population, it can be inferred that any substitute sample will have the same prerequisites for completing the evaluation, as the target population will.

In situations where the evaluation demands prerequisites only found in the target population a possible solution would be to adjust the evaluation, to avoid these demanded prerequisites. As plenty of consideration and question can be asked related to the use of substitute sampling a list of questions have been created. The content of the list is a compilation of questions derived from the knowledge obtained through literature, previous studies, and interviews. Answering any of these questions requires in-depth knowledge of the target population and the evaluation. The list is given on the next page.

- Does the task in the evaluation demand any prerequisites that only the target population have?
 - **If yes**, is it possible to identify another population with these prerequisites?
 - **If no**, does every single population have the same prerequisites?
- Does the tasks in the evaluation require any specific knowledge or experience?
 - **If yes**, is it possible to identify others with this knowledge or experience?
 - **If no**, is it possible to pass on information, knowledge, experience or/and context without introducing further biases?
- What is the overall goal of the study? What objectives are ought to be answered? What is the goal of the individual task/question in the evaluation?
 - **Usability testing** Very dependent on what kind of functionality being tested.
 - **Validation** Dependent on the change that is being validated.
 - **Value** No other sample than one from the target population can tell the value the product will bring to them.
- Is it possible to obtain the same findings from an expert evaluation?
 - **If yes**, identify what method of expert evaluation that fits the goal of the evaluation best.
 - **If no**, consider if an expert evaluation can cover some areas and compliment a user evaluation, thus saving some resources

4.2 How to proceed from here?

The questions of what to do, when the product is not suitable for user exposure were answered. The same was the question of how quantitative and qualitative approaches affect sampling.

The last question on how to approach sampling, when the target population is not available is however not answered at a satisfying level. The answer for how to approach sampling, when the target population is inaccessible further lead to a list of questions. It is not sufficient to answer a question with more questions. To answer the list of question and to obtain an answer to how substitute sampling would function in practice a method should be developed. Due to the unsatisfying answer, the development of this method will be the goal of the thesis. The method is developed for identifying when a substitute is possible, what this substitute should be and how the substitute impacts the results. For that to succeed, all aspects regarding target population, type of evaluation and, the goal of evaluation have to be part of that process. After establishing this method, the final goal will be to validate and verify the use of it. For the method to be successful, it has to be rigorous and accurate.

Chapter 5

Designing the Method for Substitute Sampling

The method for identifying when substitute sample is available will rely on analytically methods. The first objective of the method is to analyze the target population. Thorough analysis, the target population's characteristics should be mapped out. The mapping allows researchers for knowing the 'identity' of the users. Knowing the users should enable an understanding of the identity of the users and their perceptual and cognitive prerequisites. This first step of the method is implemented, as everything up until now suggests that knowledge of the users are important in every aspect of sampling.

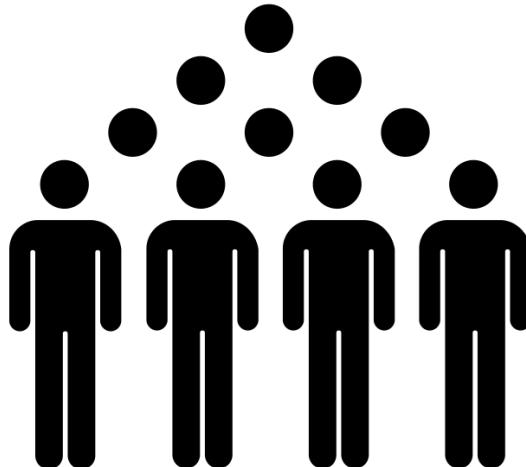
The next step of the method will include analysis of the evaluation in scope. For this method to work, the evaluation will have to be designed before picking a sample due to the goal of evaluation co-deciding what kind of sample to use. The analysis of the evaluation will focus to identify overall goals, the objectives of the evaluation and finally a thorough review of tasks. The review of tasks is conducted to establish the needed prerequisites for completing the task as intended. For this to work, the tasks will have to be condensed into operational goals.

Finally, the two analyses are compared to investigate whether the operational demands for the tasks coincide with the characteristics of the population. If the prerequisites for the tasks does not coincide with the characteristics of the target population, the possibility of using a substitute sample is enabled, due to tasks not being specifically related to the target population. Through comparison of the outcome of the two analyses, it is possible to identify the limits and advantages of using a substitute sample. A third methodological approach is applied to ensure a rigorous comparison and use of the two analyses.

When taking these steps, it should be able to argue whether or not a substitute sample is adequate. If a substitute is adequate, the results of the method should also have established the selection criteria, for picking this substitute sample. Tasks demanding different prerequisites could imply that the evaluation might benefit from being conducted on two substitute samples each representing a part of the characteristics of the target population.

In the following sections the methodology applied for the analyses is detailedly described. Furthermore, the argument for applying these exact methods are provided.

5.1 Analysis of the Target Population



In the analysis of the target population, the goal is to get to know the users. Knowing the users allows for a collection of attributes that characterizes the particular population. Through the interview with Tina, existing methods, such as personas, user profiles and user segmentation's, was suggested to be helpful for obtaining this knowledge. The ultimate method to apply for this step of the method would be personas. As personas previously have been used in for implementing user needs is deemed advantageous to apply. Furthermore, personas have in some cases be used for surrogate testing, which to the author's knowledge is similar to the idea of substitute sampling (LeRouge and Ma, 2010, p. 255)

The typical use of personas is when designing for a target population. The persona reflects the needs, skills, behaviors, attitudes and, goals the target population has. These reflections are thought into the design of the product. Unfortunately, one of the initial steps in the development of personas is to conduct or use prior research with the target population. Information on the target population explicitly demands access to the target population or previous research conducted with the specific population (Cooper et al., 2007; Junior and Filgueiras, 2005).

The demand for knowledge of the target population is not perceived as a major issue. Multiple sources of information are available in most cases, such as articles. While businesses with nearly unlimited resources might be able to obtain the knowledge themselves, smaller business is most likely to rely on previous research. Common for both small and large business' is that if the population is impossible to reach, the personas might not be as detailed as wanted. The supreme persona would be of such detail, that it covers all aspects of the target population.

Personas with lower detail are however still deemed useful, in the case of identifying substitutes for sampling. The supreme persona will most likely contain a lot of details not necessary for the evaluation or the sample. Some of the characteristics and prerequisites can, therefore, be disregarded due to them being irrelevant. For the use of personas, it is essential to keep in mind that the goal is not to find populations that fit the specific persona. The goal is to identify populations has the relevant characteristics, attributes, and prerequisites for the evaluation in scope. While the comparison of population and evaluation is described later, understanding the area of use for the personas are deemed necessary.

Obtaining a low detailed persona should not give worries. The development of a perfect persona can take very long long time. In most cases, the resources and time required for a persona will not even be available. Limitations of resources and time require some compromises in the development of personas. The compromises will most likely remove some of the details. A compromise one could imagine, would be a delimitation of the method. The delimitation would most likely cause a more general view of the target population.

The delimitation would result in more general descriptions of the population's prerequisites. The prerequisites will consist of skills, experiences, and behavior. The delimited personas are condensed into a list of prerequisites. The list shows the prerequisites that separate the target population from anyone else. Only noting unique prerequisites implies that general prerequisites for multiple populations are not to be noted. Having this set of prerequisites enables a comparison. What the population excels at and the unique skills they possess can be compared up against the actual need for these in the evaluation. Getting this far requires knowledge of how to develop personas or at least the delimited personas.

Creating a Persona

The high detail of personas and the required resources for developing it is unfortunately not available for this project. Due to this, the continuation of this project will use the delimited persona and hopefully show that this limitation of method still functions as a mean for identifying the substitute sample. The full procedure of developing a persona is however given in Appendix F - on page 161.

While the delimited persona still requires knowledge of the target population, the best advice would be to ask anyone whose knowledge of the target population exceeds and adds to the knowledge available for the researcher. Additionally, the more common behaviors and prerequisites of the target population has a chance of being identified through literature.

In the development of personas, one element is to identify behavioral variables. This step of the procedure still seems very beneficial for identifying the prerequisites of the target population.

The identification still relies on previous research and knowledge of the target population. Obtaining knowledge of the variables, requires reviews of observations, quotes and general knowledge of the target population. These reviews are used to identify what variables that identify and separates the target population from other populations. In the original method for developing personas Cooper et al. (2007), breaks down the types of variables into five categories. These are:

- **Activities** - What the user does; frequency and volume.
- **Attitudes** - How the user thinks about the product domain and technology.
- **Aptitudes** - What education and training the user has; capability to learn.
- **Motivations** - Why the user is engaged in the product domain.
- **Skills** - User capabilities related to the product domain and technology.

Through consideration of different types of behavioral variables, this step should produce several variables. It might seem like a highly detailed persona, but compared to the theory of the construction of personas, this is much more simple. Even though the number of variables is high, only a selection might prove useful in the identification of a substitute sample. Using only a selection of the variables is mainly argued with the importance of remembering that type of evaluation can make some of them inconsiderable. Variables not used could be due to their low impact on the task or questions in the specific evaluation. While trying to obtain attributes and characteristics of the users, it is essential to keep the overall goal of the analysis in mind. The outcome of identifying the behavioral variables will not be enough for developing a full persona. It is, however, the initial step, which allows for further development if resources allow.

Using only the initial step of persona development keeps the complexity of the method to a manageable level. Having the behavioral variables and thereby also the prerequisites for the target population enables the continuation of identifying substitute samples.

5.2 Analyzing the Evaluation through Hierarchical Task Analysis



Much like the analysis of the target population and the identification of prerequisites the intended evaluation has to undergo analysis. The reason for analyzing the type of evaluation is to identify all reasons for using a substitute sample. Additionally, it is to identify if there are certain pitfalls related to the sample. While the goal of other methods is to identify usability problems and save the resources it takes to evaluate with end-users, the goal here is not to identify usability problems without the use of end-users. Using the method does however not disable the identification of usability problems. This is seen as an additional outcome very beneficial for the development of a product. Instead, the goal is to identify which questions and areas of the evaluation that is only answerable for the target population, and which a substitute sample will be able to answer. To obtain an analysis of the evaluation, the *Hierarchical Task Analysis* (HTA) is chosen. The choice of method is based on Stanton et al. (2017), as he describes the method as easy to use, flexible and accurate. Additionally, the method is generic and applicable in any domain (Stanton et al., 2017, pp. 46-49). Other methods to obtain the goal exists, but requires higher knowledge and domain expertise and are often resource intensive compared to the HTA.

The HTA is a useful tool for predicting usability problems without including users. Furthermore, the HTA can be useful for identifying the operational demands set for the participants. It is still imperative to consider the burden of proof and 'importance' of the evaluation, but for now, the focus is the content of the intended evaluation.

Hierarchical Task Analysis

The typical usage of HTA is to predict design-induced errors through task analysis. The use of HTA in this context is to predict operations required for doing the intended interaction. Some alterations of the method and the taxonomy are necessary, as the method originally only identifies the operations required. It does not identify what the operations require from the participants. The only consideration to keep in mind is that the outcome should support the identification of a substitute sample. Following the procedure summarized below, it should be possible to conduct the HTA without further changes. A more detailed review of the original method of HTA is given in Appendix F - on page 167. This appendix also depicts the common ways to visualize outcome from an HTA.

Using the Hierarchical Task Analysis

The procedure of the task analysis is divided into steps. Some initial preparation is however necessary. The preparations consist of determining the overall goal of the evaluation. Furthermore, it is to map of external influences, e.g., type of technology and task constraints. The least resource intensive way of doing this is to rely on expert analyses or previous observational data that shows how technology or task constraints influence the use. If it is not possible to rely on this, other ways have to be used for the mapping of external influences.

After these initial steps, the development of the HTA is possible. The overall goal of each task has to be defined, e.g. 'Inject 20 units of slow-acting insulin into the injection-pad, using the prototype device'. When a goal has been defined, all sub-goals of the task has to be defined. Each sub-goal should be meaningful, and the combination of the sub-goals should contain all necessary interactions for achieving the overall goal defined in the previous step.

Following the example used before, the sub-goals could be 'Prepare the prototype', 'Dial 20 units', 'Inject', 'Hold until the device has dispensed the full dose' and 'Pack away device'. After defining the sub-goals, a decomposition of each sub-goal is made. The decomposition happens by breaking down each sub-goal until an appropriate operation appears. Everything before an appropriate operation is a set of goals and a description of what needs to be done.

Operations are defined as actions required to be done by an agent, to achieve the associated goal. Taking the 'inject' sub-goal could have operations like 'Check for a correct set amount of units', 'make sure injection site is prepared' and 'check state of the needle'. Decomposing all sub-goals into meaningful operations enables the development of plans. The plans typically state the order of completion of the single task and the affiliated sub-goals.

Finally, it is possible to develop a plan for achieving the overall goal by piecing together the plans of the single tasks. The plans can be both simple and sophisticated, and the complexity of the plans is coherent with the complexity of the task. Stanton et al. (2017) created a table, giving an overview of different types of plans. Table 5.1 contains the different types.

Plan	Example
Linear	Do 1 then 2 then 3
Non-linear	Do 1, 2 and 3 in any order
Simultaneus	Do 1, then 2 and 3 at the same time
Branching	Do 1, if X present do 2 then 3, if X is not present then DON'T
Cyclical	Do 1 then 2 then 3 and repeat until X
Selection	Do 1 then 2 or 3

Table 5.1: Different types of plans for the Hierarchical Task Analysis (Stanton et al., 2017, p. 50)

The desired outcome

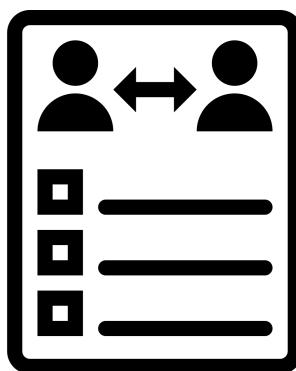
The outcome from the HTA is thought to be supportive, due to the insights it provides for each task. It still, however, does not include a prediction of human-related prerequisites required for doing the operation. Prerequisites have to be assigned to each operation to accommodate for the missing human relations. As the operation is a required interaction, it is possible to assess how and which prerequisites that are necessary or influences the interaction. This addition to the method seems best applicable after conducting the step of the *plan analysis*.

Due to the addition of a step and the superficial walk-through of the method, the listing below shows the required steps and sequence for using the Hierarchical Task Analysis. The step marked in bold is the added step. All other steps are the original steps from the method. As aforementioned, greater detail of each step is accessible in the appendix.

1. Defining tasks under analysis
2. Data Collection Process
3. Determine overall goal of tasks
4. Determine the sub-goals of the tasks
5. Sub-goal Decomposition
6. Plans Analysis
7. **Prerequisite Attribute Prediction**

Completing the task analysis should result in vast knowledge of the evaluation and the demands each task sets for the participants. Knowing what is demanded by the participant and knowing the participant's general prerequisites allows for comparison. Additionally, it allows researchers to make informed predictions of what errors that are induced by multiple elements. These elements consist of the tasks, the prototype being tested and the type of sample.

5.3 Comparing the Target Population and Task Analysis



After obtaining an analysis of the target population and the evaluation, a comparison is possible. Each operation identified through the decomposed sub-goals should be held up against prerequisites for the target population. If the operation has no specific requirements for the interaction, it suggests that the target population is not needed. If it does have specific requirements related to the target population, it is possible to take the required prerequisites and check if another population fulfills this requirement. If other populations do, they possibly have a chance of functioning as a substitute.¹ Settling the required skills and knowledge for a specific operation is necessary for identifying the substitute population. E.g., if it is user experience with injectable subcutaneous medicine, another population with a disease treated with such medicine might be adequate for substitution. These considerations should be taken for each of the operations listed in the sub-goals. If 9 out of 10 operations requires specific characteristics, the argument for using a substitute sample, might not be strong enough, as the demand for the target population is evident.

To compare and predict design- and sample induced errors a third methodology is applied. The hope is that the applied method raises the robustness of the comparison. The methodology is deemed needed as predictions usually are profoundly influenced by subjective opinions and ideologies.

SHERPA

To obtain a structured and rigorous comparison of the two analyses some methodology is needed. Because a task analysis has been conducted the most advantageous method to apply will be a method benefiting from this. Therefore, a reasonable method to apply would be the *Systematic Human Error Reduction and Prediction Approach* (SHERPA) (Stanton et al., 2017). SHERPA is an extension of the HTA, and the initial step of SHERPA is to use the HTA as input. While the focus of the method is to predict potential human errors, some modifications to the

method are necessary for enabling identification of substitute sampling. The identification should predict when, how and why a substitute sample is available. Obtaining this knowledge additionally identifies the limits of using a substitute sample.

Through further analysis of the HTA, the SHERPA applies some structure for the comparison of evaluation demands and participant prerequisites. It classifies the tasks through predefined taxonomies. The classification then enables the use of an error mode taxonomy that adds some robustness to the prediction of design-induced errors. The errors are further described and so is the consequences of the errors. The method further analyzes the tasks, which in the end enables the researcher to establish some consequences of not using the target population in a sample. Knowing these consequences then allows for assessment of how critical it is to use the target population. If it is deemed critical to have a sample of the target population a substitute sample will not be wise to use.

How to Use SHERPA

Using SHERPA requires a sequence of steps. These steps are superficially described in the below. A thoroughly described use of SHERPA can be found in Appendix F - on page 170. In the appendix, it is however only the steps of the original method that are described. The adjustments needed can only be found in the main report.

The goal of SHERPA is to classify each task. The classification of tasks happens by following the taxonomy of the method. These classification taxonomies are:

- Action (e.g., pressing a button, pulling a switch, opening a door)
- Retrieval (e.g., getting information from a screen or manual)
- Checking (e.g., conducting a procedural check)
- Selection, e.g., choosing one alternative over another)
- Information Communication (e.g., talking to another party).

After this, credible error modes associated with the classification of tasks are assigned, and a short description of the potential error written. The description relates to the assigned taxonomy of SHERPA and expertise within the domain. Next is the analysis of consequences. After the consequences is an analysis of potential recovery of these consequences. Finally, each potential error is assigned a probability for occurring and at last a criticality analysis is performed, to assess the influence of the error. The last step is to perform a remedy analysis, which contains proposals for error reduction strategies (Stanton et al., 2017, pp. 143-153).

Including not only tasks

Using the HTA and the knowledge gained from the analysis of the target population does not necessarily show the full picture. In most evaluation tasks will be followed up upon with a series of questions. These questions are of the same importance as the tasks. There might be questions that are unrelated to the potential substitute sample. The participants of the substitute sample will not be able to give a valuable answer if they cannot relate to the question.

Furthermore, through the interviews, it was discovered that information given throughout an evaluation could have an impact. Any information given throughout an evaluation might change the way a participant answers. Telling a diabetic patient that the prototype device is to be used to inject three times a day might not surprise the participant. If the participant does not have diabetes, this might surprise them - or even fear the participant. These kind of influences are hard to prove, and the amount of impact they have is not known. Any irregular or unusual behavior influenced by information should be accounted for if using a substitute sample.

Desired outcome

SHERPA relies on the task analysis, and it seems useful for achieving informed predictions of potential errors. The original use of SHERPA is intended for identifying substitute samples, and therefore two additional steps are added.

The first step is another consequence analysis, which the scope of identifying what consequences it will have to use a substitute sample compared to a sample from the target population. For identifying the consequences, it is suggested to rely on the knowledge obtained from the original steps of SHERPA. Reviewing the predicted errors and consequences for each sub-goal of the tasks can be related to the potential substitute sample. If additional unique errors are induced by using a substitute sample, then the consequence should describe these.

The second step is then to assess how critical these errors are. This assessment is suggested to rely on a comparison of the consequences just identified and the primary objectives of the task. If the consequences do not disable the substitute participant to fulfill the primary goal of the task, it could indicate that a substitute sample could be adequate to the target population. Finally, if the criticality of using a substitute sample is assessed high and the target population is impossible to reach options remain.

An option could be the review the consequences of using a substitute sample. The possibility of retaining the opportunity for a substitute sample can be maintained if there is a possibility to adjust the task in question to avoid the consequence. This procedure is however only possible if the adjustments to the evaluation do not hinder the researchers to collect the intended data of the task.

Another option is to review the consequence and criticality. Imagine a task with the only requirement being that the participant has experience with taking injectable medicine. The consequence of not using a sample with this experience is deemed very critical, but all other aspects are unnecessary for this particular task. This result allows for setting a sampling frame targeting all people with this kind of experience, and not only people with diabetes. Not relying solely on people with diabetes allows for a broader sampling-frame which enables the use of a substitute sample. The below list shows the full procedure of performing a SHERPA and includes the added steps. The added steps are marked in bold. The remainder of steps is from the original method.

1. Hierarchical Task Analysis
2. Task Classification
3. Human Error Identification
4. Consequence Analysis
5. Recovery Analysis
6. Ordinal Probability Analysis
7. Criticality Analysis
8. **Sample Consequence Analysis**
9. **Sample Criticality Analysis**

Due to the added steps being a suggestion, there is, unfortunately, no well-established taxonomy for the aspects regarding sampling. To obtain a taxonomy for the sampling aspects of the method many iterations of the method is needed. As resources and time do not allow for iterations, there is no goal of establishing this new taxonomy.

A finished SHERPA would enable the possibility to take an informed decision regarding sample. This is true no matter if the outcome shows the need for the target population or shows an option for a substitute. Up until now, there have not been a single end-user included. There are however still aspects to consider, as the use of a substitute sample requires attention in both the conduction and analysis of the evaluation.

5.4 The Full Method

The SHERPA helps to identify substitute samples, where it is applicable. When the procedure is carried out, it is possible to conduct the chosen evaluation. Multiple samples each with a characteristic they share with the target population can together represent the full specter of the target population, which requires a more diverse sample. This diversity in sample introduces some uncertainty about representativeness, as the variance of the sample might be high. Higher variances are only speculations, but a larger sample could reduce the uncertainties introduced by the mismatch in the sample versus target population. This compensation would also allow for the results to have higher accuracy, according to Kothari (2004).

For the conduction of evaluation, it is important to conduct it with respect to the chosen method, as always. The usual rules for conducting research is also valid when using a substitute sample. These rules usually imply sample size and context. The same applies to the analysis. It is, however, important to pay respect to the analysis, in a new way. Using a substitute sample requires all interpretations to be done carefully, as it is a substitute sample and not a sample from a target population. Participants from the substitute sample will not be able to represent and articulate how a product brings value. The aspect of value is argued by the research conducted by Jensen (2017), showing that the user experience for a person with diabetes includes aspects based on their disease. There is no evidence that the user experience for the substitute sample also is valid and generalizable for the target population. With these thoughts in mind, it should be possible to complete a user evaluation fully. In Figure 5.1 is the complete procedure.

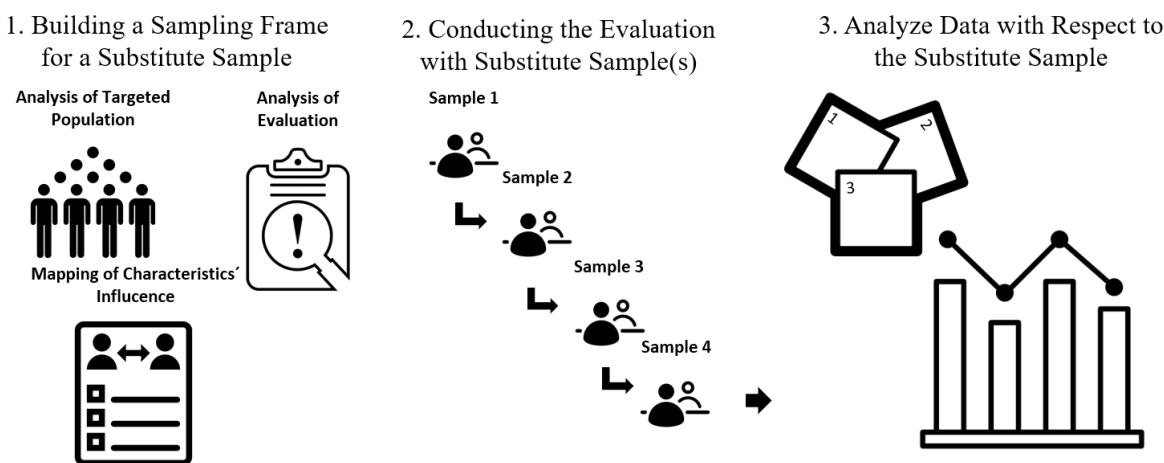


Figure 5.1: Depiction of the full procedure. Each square represents a step in the method. The first step consists of the three analyses. The order of the analysis of the target population and evaluation are indifferent. It is however not possible to map the characteristics' influence in SHERPA without knowledge of both the target population and the HTA.

While the first step focuses on the possibility of using a substitute sample, they also provide additional input. The original intended use of both personas, HTA and SHERPA, is still preserved as there have only been additions to the methods.

The use of personas can, if fully developed, be used in the phases of design and as an argument for design changes. The HTA and SHERPA also still provides the possibility to identify and predict usability problems without conducting the actual study. The original elements of the methods also serve as a tool for argumentation in design, development of prototypes, and, evaluations.

The development of the method is purely based on theory, literature and, empirical data and there is no evidence that it functions as intended. Therefore, the next step is to validate the method.

Chapter 6

Validating the Method

Validating the developed method can be done in several ways. At first, it was chosen to replicate a previous study from Novo Nordisk. A replica study allowed for comparison between the use of a sample from a target population and the sample identified through the method. To replicate the study, a lot of resources and time went into the planning of the several aspects customarily considered when designing a study. One major problem did, however, hinder this procedure. The stimuli used in the study were meant to be used in the replica of the study. Unfortunately, late it was discovered that some of the electronics used for controlling a wireframe, had taken severe damage from the transport from the last study. It was deemed possible time-wise to fix these problems and order new components for the study to be conducted. This assumption was not right which excluded the use of this particular study. It was predicted, due to the time and resources wasted on the planning that time did not allow for replication of another study. Thus another procedure was chosen for validating the method.

It was thought to do the full procedure. The full procedure implied analyzing the target population, analyzing the evaluation and finally mapping the characteristic's influence on each task. After finishing the analyses, the study should have been conducted, with a sample taken from a substitute population.

Instead, the validation of the method will be done through a workshop. The goal of the workshop is to observe the method in the hands of employees at Novo Nordisk, and see if the method allows for reaching the intended goal. The workshop-approach allows for feedback. Thus optimization of the procedure will be a possibility. Instead of validating the method result-wise it is instead validated use- and functionality-wise. The workshop-format will not allow for validating the aspects of conducting and analyzing a study where a substitute sample has been drawn, based on the results from the analyses. Instead, the workshop will revolve around the analysis of population, the use of HTA and the use of SHERPA. Through the use of the method, it is also wanted to obtain general feedback useful for the optimization of the procedures.

6.1 Design of Workshop

After the planning of the workshop (setting a date and inviting participants) the actual design of the workshop was begun. The workshop consists of five different steps that are performed in sequence. The list below shows the steps and the sequence. Text in *italic* was not shown to the participants, but merely function as an overview of the type of assignment, for the reader. The five steps are:

1. Introduction
2. Assignment 1 - *Obtaining attributes*
3. Assignment 2 - *Task Analysis on a case*
4. Assignment 3 - *SHERPA on the Task Analysis*
5. Discussion and Feedback

Each of the steps is divided into multiple stages, to get some structure to the workshop. These stages will be described further, as these are the content of the workshop. All choices regarding the structure and information provided throughout the workshop are further argued in each of these descriptions. The full presentation (all slides) used in the workshop is given in Appendix H - *The Workshop and The Tools for Conduction* on page 180. Immediately following the actual slides used in the presentation, are all presenters notes. These supported everything not appearing in the slides, such as facilitator notes and actions to be taken during each stage.

Introduction

The purpose of having an introduction is to give a summary of what I am investigating. It is not a complete presentation of the whole thesis, but it is to remind what I am working with in general. Communicating the problems within sampling should be enough information for the participants. Additionally, it contains some of the more practical information, such as an agenda. The introduction is further divided into four stages.

At first, a summary of the project itself is given. The summary is done, to give some context to why we are conducting the workshop and to present the overall goal. It is essential to mention, that none of the information given is specific, as it is wanted to present the participants of the workshop with as little information as possible. This is argued with that if I presented the whole content of the workshop immediately the participant's thoughts would not be solely on the single assignment when they have to participate actively.

After this initial presentation, the agenda is presented. The agenda follows the list given above but also includes feedback sessions and short breaks in between each assignment. The feedback sessions are placed at the end of each assignment to get feedback on the individual assignment and the used method. It is predicted that if the feedback sessions were at the end of the workshop, some of the essential and more specific feedback would be lost.

After the agenda, a bit more specific information is given. In the beginning, the problem of sampling was told, and in short, the method designed for solving these issues is now given. The short introduction introduces what the participants are going to work with. There is still no mentioning of the actual assignment as personas, tasks analysis or SHERPA. Instead, it is described as analyses of population and evaluation and final prediction of errors.

For the final step of the introduction, it is deemed very necessary to describe the area of use for this method. All participants are experts in UX, usability and, design and are used to give input to product development. Additionally, some of the participants are very familiar with the processes of sampling and recruiting. It is imperative that all participants be on the same side and understand that this procedure applies to situations where the sample is impossible or hard to reach. It is not a suggestion for substituting samples in all cases of evaluation. Furthermore, there will in most cases be limits to what a substitute sample can provide, whereas everybody has to acknowledge that the use of a substitute sample is very dependent on the goal of evaluation and the expected outcome.

Assignment 1

In the first assignment, the goal is for the participants to structure a set of prerequisites that characterizes people with diabetes. To apply structure to the process, inspiration from Spool (2004) have been used, who was further inspired by Kawakita (1975). The full description of the applied methodology is found in Appendix I - *The KJ-Technique* on page 219. The reasoning for including some methodology to the first assignment of the workshop is, to have all the participant collaborate and achieve an established set of prerequisites. Besides, the KJ method further categorizes and ranks these. Maybe these categorizations of characteristics and attributes can be the foundation of some of the taxonomy that is lacking sampling-wise in the current state of the methods. In addition to the wish of obtaining some initial taxonomy for sampling, KJ excels at other things. It has been mentioned before, but sampling seems to rely on subjective opinions. The KJ-technique excels in obtaining an objective group consensus from a variety of subjective and opinionated data, which might counter the subjective approaches and thoughts on sampling (Spool, 2004).

As with the introduction, the first assignment is divided into stages. At first, the problem is further described. The further description is done to make the participants focus on the specific assignment. The focus is hopefully achieved by telling the participants, that any given population has some set of attributes and characteristic that ultimately describes and identifies how the population in scope is any different from all other population. The next thing given is a plan for obtaining these prerequisites. This implies an introduction to the use of the method, which is the previous described delimited method of personas. In this step, a short description of what personas are is given. Additionally, all knowledge needed for identifying prerequisites is provided as this is needed in the assignment, e.g., the different types of behavioral variables.

Finally, the actual assignment is presented. It states that the goal is to establish a set of attributes and characteristics of diabetics. It follows the structure of the KJ-technique and the predefined steps of this method. Each step is introduced separately for the participants not to get ahead and to focus on the current step. The central question they are trying to answer throughout assignment 1 is *What attributes, and characteristics describe people with diabetes and helps to distinguish the population from other populations?*

The last stage of the first assignment is feedback to the method the participants just went through. Did they find anything confusing, something missing or something being too much? Do they have suggestions for other implementations of the method? Questions like these. Most importantly, the participants also have a chance to ask questions that might have developed throughout the assignment. If any of the upcoming assignments answers the questions, the participant must kindly wait to get the answer. This way focus is kept on the actual assignment and not on anything unnecessary.

Assignment 2

Presentation-wise a similar structure for assignment 2 is chosen. The similar structure leads to four stages. These consist of a presentation of the problem, a plan for solving it and finally the assignment the participants are going to solve. After solving the assignment, a feedback session is planned. Presenting the problem implies informing the participants about how answers to tasks and questions might differ and rely on the type of sample being used. Informing about other aspects will hopefully expand the mindset of the participants and start thinking about how the evaluation influences the choice of sample.

Next, a plan for solving or at least identifying how the type of tasks influences the outcome based on the type of sample. In short, it is an introduction to the hierarchical task analysis, which many of the participants are already somewhat familiar with. Introduced shortly, is the added layer consisting of assessing the required prerequisites. No more information is given, as it is deemed unnecessary and time-consuming. As they have been introduced to the problem and the suggested solution it is possible to start the assignment. Even though most participants will have experience with task analysis, a short familiarization is conducted.

In the familiarization, all participants are asked to do an individual task analysis following the procedure described by the method of Hierarchical Task Analysis. All participant have the same case, which is to do a task analysis on the task "Make a cup of tea." There could have been multiple cases, but it was decided only to use one. The familiarization has a hidden agenda; it allows to show how subjective a task analysis is to the participants. Additionally, the choice of using only one case and asking the participants to do the analysis separately allows for comparing the outcome. Since subjective opinions and expertise profoundly influence the HTA, it is expected that some variance will be present in the outcomes. This check of variance cannot be done statistically. Thus it will be a subjective measure.

It was considered using multiple cases, which in contrast to the check of variance would provide a measure for the robustness of this step of the method. The argument would be, that if the method is applicable in multiple situations and further tested it could be deemed generic. Multiple cases are however not chosen. The reason for not choosing multiple studies is that it requires more resources to facilitate multiple cases at once. Additionally, it seemed more interesting to check for variance as this result could be more generic and generalizable when only having one iteration.

After completing the familiarization, the participants are asked to collaborate and make a task analysis. The case used in the assignment is a previous study conducted in Novo Nordisk where the sample was people with diabetes. The study has been subjectively chosen based on the diversity of the tasks. It contains simple handling tasks, but also tasks requiring specific knowledge. The assignments are: (1) Attach Device (2) Inject (correct amount) (3) Inject (wrong amount) (4) Detach Device. Correct and wrong amount refers to the number of units taken compared to what is shown in a wire frame of an app. This were to observe participants reaction to inaccuracy. In Appendix H - *The Workshop and The Tools for Conduction* on page 217 is the full case they had to work with.

It was considered to have the participants make each their task analysis. At the end of the assignment, the participants would have to agree on only one of these to proceed with. Only proceeding with one was not chosen as it broke the natural flow of the workshop. Additionally, a task analysis would probably never be conducted individually, without input from others. Lastly, it is cumbersome to ask for individual analyses, when only one had to be used further on. It seemed like the participants would waste their time, which is never a good thing. While one could argue that they all could continue to work individually, this seemed dumb for several reasons. One reason is that the next assignment could benefit from being done in a group. As the last step of the second assignment, all participants are asked to provide feedback to the assignment, the method and the format of the workshop itself.

Assignment 3

Identical to the previous two assignments the overall problem is presented. The problem is told to be that the two previous analyses still does not account for the sample itself but rather functions as two segregated analyses. Similarly, the participants are then told, that a possible solution to this is to use the results obtained from the previous assignments, to predict human errors with regards to the single task. The comparison and use of previous results allow us to convert the results from the two segregated analyses into a tool for argumentation in product design, evaluation, and type of sample. Ultimately, this tool allows for making informed decisions.

The SHERPA is then conducted step by step in a group with all participants. The argument for doing it in groups is the same as the one used for the participants to do the HTA in collaboration. Additionally, many of the things in SHERPA is subjectively assessed, and ongoing discussion can probably result in a more objective and concise outcome. The order in which they perform the SHERPA is also the same as listed in Chapter 5 - *Designing the Method for Substitute Sampling*.

After having conducted the SHERPA, the associated feedback session is done. Once again, they are merely asked to give feedback to the assignment, the method and the format of the workshop itself.

Discussion and Feedback

After having conducted the three assignments, approximately 2.5 hours have gone. It is however wanted to clarify how it is thought to proceed from here. Having a finished set of attributes, an HTA and a SHERPA does not give any actual sense if there is no understanding of what to do next. Therefore, it is explained that the tool is useful for identifying **if** it is possible to sample from an alternative population. If this question can be answered with a simple 'yes' all information in the SHERPA should be able to pinpoint what criteria the sample framework then should contain. E.g., if the only requirement is experience and use of subcutaneous injections, it is possible to sample from any injection treated diseases instead of focusing solely on the diabetics.

If time allows, it would also be interesting to discuss some of the disputable aspects of the literature. The disputable subjects could be a discussion of subjects as how to ensure that context and participant motivation is the same in the test scenario as in real life.

The discussion is left open as there might also be many questions. Finally, all participants are asked to write two Post It's. The first one is answering "I liked" and the second is answering "I wished". Answers to these questions help to identify what they liked about the workshop and what they wished for the 'next' time. This format of feedback is not used for the feedback sessions after each assignment, as more specific feedback to the individual method is wanted. The more specific feedback implies if they learned something, if they find the use of the method appropriate and if they have any suggestions for adjustment or optimization.

6.2 Data Collection Method

For the collection of data, a structured procedure for all three assignments applies. The participants are, through the workshop, going to produce a variety of data, which they are supposed to put on Post It's. Therefore, everything the participants are to note down for each assignment and the stages thereof is structured. Each step in an assignment is color-coded, meaning that the specific step uses one color of Post It's. In some cases, where colors are to be used multiple times, each post it is also given an identifier.

This structured procedure help to backtrack, if anything exciting shows under the conduction of the workshop. It also allows for an overview and possibly further qualitative analysis.

There are however more data than the data belonging to the methods. All feedback is noted down, as this is what helps validate the use of the method. Additionally, the content of discussions (both ongoing in the assignments and discussion of sampling in general) is also noted down. This type of data is to be treated more subjectively, due to it being purely based on articulations and observations.

Both the workshop and the collection of data have been designed and is ready to be conducted. The recruitment of participants and booking of facilities have been done in parallel to the designing the workshop.

Chapter 7

Conduction and Review of The Workshop

The workshop was conducted May 17th at Novo Nordisk facilities in Hillerød. Six participants took part in the workshop, and a single participant was scheduled to participate later. Unfortunately, one participant was forced to prioritize other projects above participation in the workshop. The missing participant caused the workshop only to have five participants, and the participant scheduled to appear later did not show. Additionally, two of the remaining participants had to be elsewhere half an hour before the end of the workshop. The juggling of participants resulted in a workshop with five participants in the beginning and was further reduced to 3 participants in the end. All participants had professional backgrounds in Usability, User Research, Design or Industrial Design.

It was not optimal to run the end of the workshop with such low numbers of participants. With two weeks notice it can be hard to find participants who can prioritize three hours of their time over regular assignments in their everyday work. The workshop was conducted, even though the number of participants could have been higher. It produced feedback for the methods and the workshop overall. Additionally, it produced results of using the methods.

In the following sections, all feedback and results are presented. The structure for this is first to present the more general things about the workshop, to give an overview. Next, an individual description of the three methods used in the workshop. In this description, a general review of the method is provided. Following the review, the results produced by the participants is shown and further analyzed. As the outcomes from all three assignments already have been subject to categorization, it is subjectively chosen only to discuss and interpret what the participants found. The exclusion of further analysis is done, to be true against the output from the participants.

The review, interpretation, and discussion of the workshop are going to provide input for optimization in two areas; 1) optimization of the workshop and, 2) optimization of the methods. The suggested optimization are described in context to the results, but in the end, a summary of the actions to be taken is given. These will, however, not include suggested optimization for the workshop. Instead, the summary will solely focus on the suggested changes and alterations of the methods. The optimization for the workshop is instead provided sequentially through the chapter.

7.1 The Workshop in General

The workshop took the estimated three hours to complete, but there were small problems throughout the whole workshop, which caused some ad hoc changes. One of the biggest problems was that the designed method never was intended to be performed in only 3 hours. The method did not fit the format of a workshop. Due to this, the use of the methods seemed a bit superficial, which were also commented upon by the participants. The participants understood the idea behind the implemented steps, but time did not allow for discussing and using them in detail. The lack of discussion was due to both the relatively low number of participants and the extent of the assignments. Initially, the assignments were thought as easy, but participants used too long time discussing the original steps of the method. As for the discussion part, a wish for more participants was articulated which in fairness compared to the number of participants is understandable.

Another problem that was commented and showed throughout the sessions was that the methods were 'too' academic to be implemented in everyday work routines. A wish for a more flexible structure and use of the method were present, which clearly shows a need for the method to be optimized. Furthermore, the use of the methods showed a need for more clarified definitions on the taxonomies used in HTA and SHERPA. The exact reasons for this need are given in further detail in the reviews of each assignment. At the end of the workshop, each participant was asked to answer two general questions; these were "I liked ..." and "I wished ...". The questions focused on the whole workshop and thereby everything could be answered. All answers provided by the participants is given in Table 7.1 below. All content of the table has been translated from Danish into English.

I Liked...	<ul style="list-style-type: none">... Insight in sampling... That we were allowed to try the method on a relevant case... That the structure of the method seemed innovative and useful... Learning about a new more structured approach to argue on what users to use for a given study... Being more people to discuss with. I removed my bias on how I normally do when making task analysis
I Wish...	<ul style="list-style-type: none">... It can take so many directions - something could be more specific/simplified?... That the assignments were a bit less extensive, to see the full potential of the method... That it was a little less theoretical if I have to apply it in my everyday work.

Table 7.1: Answers for "I liked..." and "I wished..." from the end of the workshop.

7.2 Creating a Foundation of Personas

The use of the KJ-technique was an excellent decision to make. Doubt was expressed by a participant, as it seemed impossible to obtain unanimously consent, without talking. Furthermore, some questions arose for the focus question. The focus question was "What attributes and characteristics describe a diabetic?". Some participants initially did not think it was possible to identify these for the whole population. They were biased by their rational thinking, that people with diabetes are divided based on either type 1 or type 2 diabetes. After completing the first assignment, some participants were impressed that the result applied to both types. Diabetes is however very cultural, meaning that the attributes and characteristics can be very different depending on nationality, e.g., people with diabetes living in Denmark and diabetics living in the USA. Culture has to be considered if real personas had to be created. Having a single persona for diabetics is not likely. It was however never envisioned that a single persona would sufficiently describe all diabetics. One could argue that a more specific focus question would have been better for the workshop format.

It was never the intention to develop complete personas through the workshop, as this requires much more time. It ought to reduce the extent of the assignment, which caused participants to question the purpose of the actual assignment. It was however explained that this should be perceived as the beginning of persona development. There was, however, a need for some ad hoc changes. Initially, participants had to vote for the three most important groups for each of the five predetermined types of behavioral variables. There were not enough groups for each type of variable. Hence a change was needed. The participants were instead asked to identify the three most important aspects overall, across the five types.

After this change, the participants completed the first assignment. The feedback they provided was that the five behavioral variables did not account for all aspects. They missed the emotional aspects a person with diabetes has, and they felt that none of the types covered user experience aspects. The most critical issue experienced was that the purpose of the assignment did not appear clear enough for the participants. The missing purpose could imply that the development of personas is difficult to reduce in complexity and still maintain the purpose of it.

Results from Analysis of Population

Through the workshop, the participants used the KJ-technique to obtain a prioritized list of attributes and characteristics describing people with diabetes. As the assignment included categorization and thematizing, there is no need for doing this in the analysis of the data. Instead, the categories and themes will be interpreted.

The results have to be presented if they are to be subject to interpretation. In Table, 7.2 on the next page a summary of the categories and their content who were deemed a priority is shown. The full list of categories and their content can be found in Appendix H.1 - *Results from the first assignment of the workshop.* on page 217.

Activities		
Categories	Score	Content
Need for care	2	Typically has a support who is indispensable in order to feel safe
Control	4	Has everything necessary for treating diabetes under control Monitors food intake Makes calculations and systems to keep track of the disease They balance many parameters (food, exercise, medicine) They keep track on different kinds of insulin
Medical treatment	6	Takes injection at meals Takes injections before going to bed Inject insulin multiple times a day They take pills every day They measure blood sugar multiple times a day They control a pump They often carry insulin pens with them

Attitudes		
Categories	Score	Content
Positive approach	1	Happy for pen-shaped device
Inconvenience	1	They find it a hassle to start on medication and they wish to stop taking it

Motivation		
Categories	Score	Content
Identity	3	They want to drop the medication (type 2) They want to forget about the illness Interested in technology, as they want to feel 'normal'
Anxiety	3	They are motivated, as they don't want to get worse Anxiety for hypos Afraid of seizures

Skills		
Categories	Score	Content
Super Users	2	They usually inject insulin in many years and become 'Super users' Injections experts Daily training with devices

Table 7.2: The categories who were prioritized by the participants. Only categories which got a score is shown.

Interpretation of Results

As the table shows, plenty of attributes and characteristics were brainstormed and added. Many of them seem to overlap or have some link between them. Taking the only '*Skills*' attribute and comparing to the rest of the attributes shows what it takes to become a 'Super User'. Additionally, it seems that some misunderstanding of the predefined types of variables has been present. Some of the items not categorized under '*Skills*' are quite explicitly an actual skill, which is especially true for the items under the type '*Activities*'. Furthermore, many of the items in activities seem to be the action for obtaining a skill, e.g., measuring blood sugar, taking pills and, controlling a pump. It is quite strange that the participants have been able to determine some actions that require some certain skills, but not necessarily have noted these skills. Using an injection as an example would require some skills and knowledge about needle handling and dose control. Finally, it shows that the participants prioritize handling and action-based attributes and characteristics. The high focus on handling aspects could be due to the background of the participants, as almost everybody had a professional background in usability and user experience evaluation. It could further be, that the participants were biased by not having a category that they felt could hold emotional and user experience aspects.

It implies that the definition and information about the five types of behavioral variables need to be even more precise. In the definition of these types, it should also be stated how to differentiate between them, as some categories seem misplaced or could belong to two different types of variables.

While the skills in Table 7.2 shows a priority made by the participants, the remaining categories and items are not to be ignored. Even though they did not get to be prioritized, they still account for what it is like to be a diabetic. If a persona had to be developed all the data would have to be used in the development. The data collected through the workshop also seems applicable for the creation of behavioral variables. Using the behavioral variables further implies taking each attribute and identifying the end-point on a scale. With, i.e., needle fear such endpoints could be low fear and high fear of needles. Next step would be to look at actual footage of people with diabetes and start placing them on these scales based on their answers and behavior.

Finally, the complexity of creating personas was said to be too difficult to reduce without introducing confusion and losing purpose. It implies that a less complex, '*discount*' version of personas might be more tangible, in situations where personas are too much.

7.3 Using Task Analysis on A Case

Using the task analysis for obtaining knowledge of a given evaluation worked as intended. While not all participants had the same experience with task analysis, they collaborated and helped each other through. Obtaining a uniform understanding required some discussion and inevitably slowed the progress. The discussion resulted that the task analysis was not complete when time ran out. Some of the discussions referred to some vague understanding of the procedure of doing a task analysis. The impact of prototype-design and functionality was also covered, and it was especially articulated that "It is the design that decides". Even though the participants did not finish, the outcome of the discussions was deemed very helpful for the participants. It is, however, not that helpful for the workshop, as the task analysis were to be used in the third assignment for predicting errors.

With hindsight, the task analysis should not have been developed by the participants. As they already had the expertise and knowledge, a complete task analysis should have been used as a case. A complete task analysis would allow the participants to focus solely on the added step, which tries to create coherence between the required operations and the required prerequisites to fulfill the task as intended. Putting a focus on the added step would have solved two problems; 1) the feedback would target the addition to the original method and, 2) a better basis for using it in SHERPA would have been established.

Results from Task Analysis Familiarization

Having the participants familiarize with task analysis initially shows that there are enormous differences. Looking further into the result shows that most of the participants had the same understanding of the task analysis, but had different strategies. E.g., participant 1 creating all overall goals first and participant 5 creating and finishing one step at the time. These can mainly be described as domain expertise, and subjective opinions regard how detailed a task analysis should be. In all fairness, it should also be mentioned, that the familiarization was very time limited (10 minutes), which also can explain some of the variances in the results from there. Additionally, the familiarization shows that domain experience matters when doing a task analysis. Participant 4 was the only participant, who do not work with usability in everyday work and this participant was the only to provide a different task analysis. In Figure, 7.1 is a depiction of the tasks analyses each participant created.

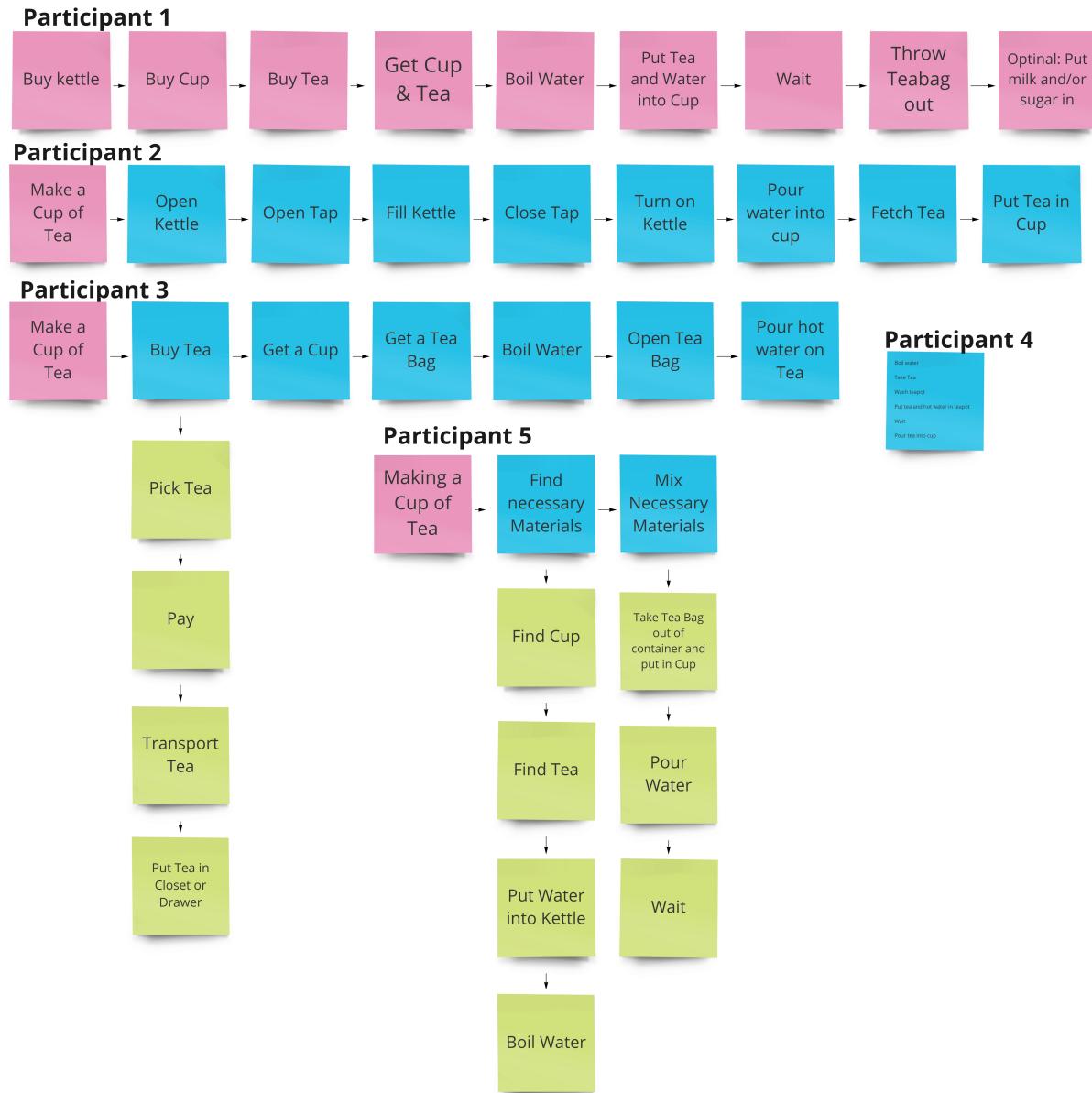


Figure 7.1: Depiction of the participants' results from the familiarization. As seen, they are different both content and detail wise. A pink Post It is overall goals, blue Post It's are sub-goals and green are operations. Participant 4 misunderstood the color-significance.

After having familiarized and all participants had seen how subjective a task analysis is, they had to collaborate. Through collaboration, they had to make a task analysis of an internal study from Novo Nordisk.

Results from Task Analysis

The participants had to work together and form a task analysis, based on a case. The presented case had four larger task and a sequence of follow-up-questions. The participants were asked to focus on the tasks and not the questions. This was to ensure that an outcome suitable for SHERPA would be produced. From the 40 minutes allocated to completing the task, the participants 'completed' two out of the four larger tasks. As aforementioned, the time allocated for the assignment did not account for the extended discussion. This discussion gave some insights, but unfortunately also hindered a fully complete task analysis. In figure 7.2 is a depiction of the task analysis the participants developed for the first task.

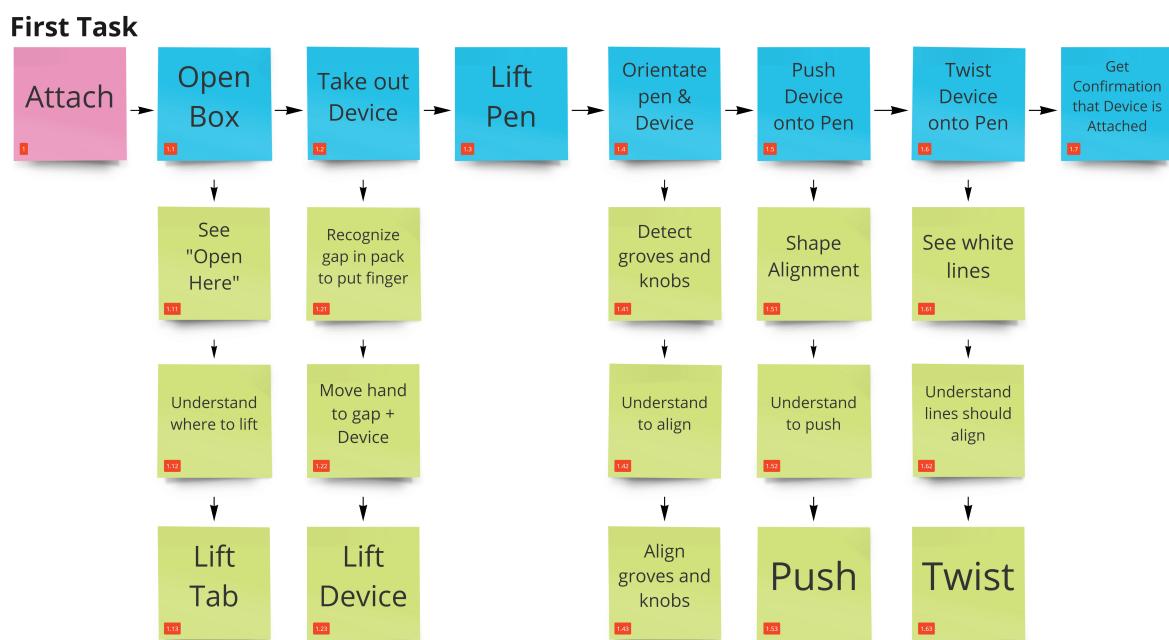


Figure 7.2: Task analysis for the task "Attach". The pink Post It is the overall goal, blue Post It's are sub-goals and finally the green Post It's are operations. This figure doesn't shows the predicted prerequisites.

As seen, the first task was deemed to have seven sub-goals and 15 operations. To keep the figure readable, the last step is not shown in the depiction. Instead, the required prerequisites are listed in Table 7.3.

For the second completed task, a similar depiction of the results is shown in figure 7.3. This task had the overall goal "Inject", which by the participants were divided into five sub-goals and ten operations split across these.

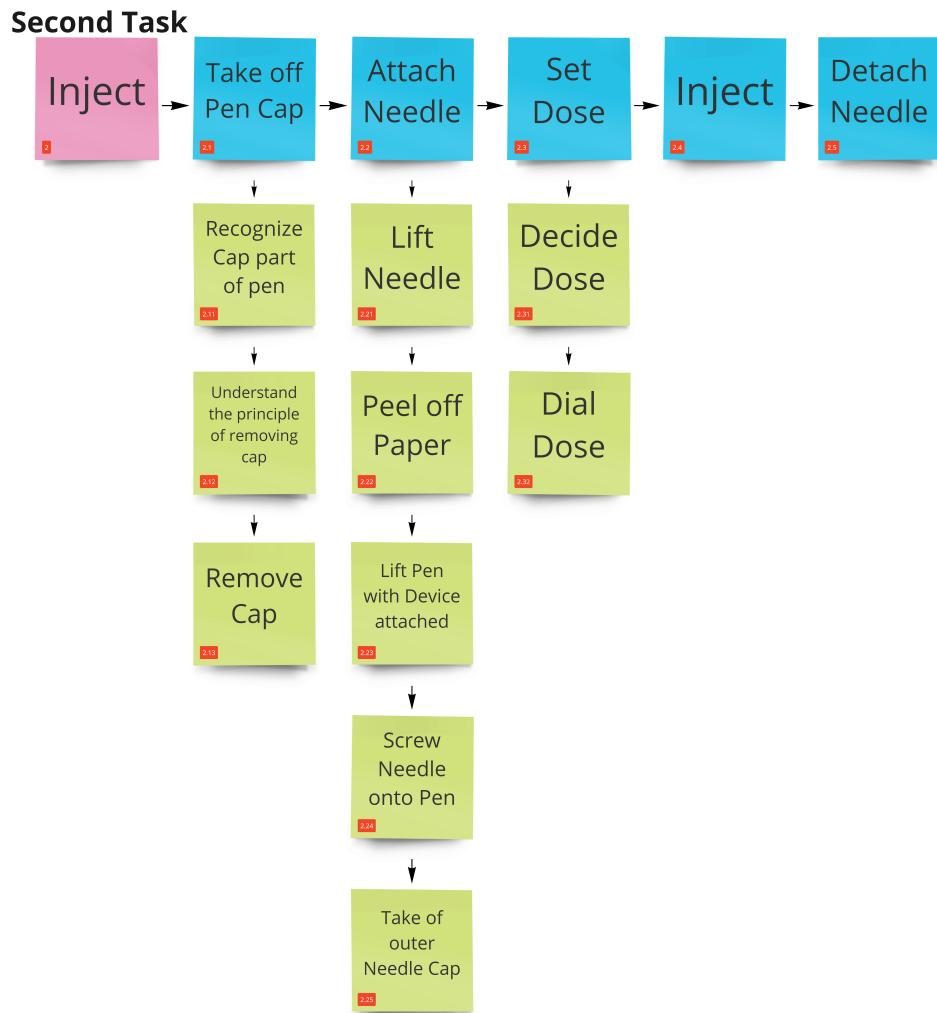


Figure 7.3: Task analysis for the task "Inject". The pink Post It is the overall goal, blue Post It's are sub-goals and finally the green Post It's are operations. This figure doesn't shows the predicted prerequisites.

The required prerequisites predicted by the participants are listed in Table 7.3. Each prerequisite refers to the tag, shown in red on each of the Post It's in figure 7.2 and 7.3. From the table, it is possible to see that some fields do not have an entry. Whether it is a mistake or that the participants did not feel that the sub-goal could be further broken down isn't known. However, compared to the results from some of the other sub-goals it could indicate that time ran out and that they did not finish. Only the tasks, which the participants of the workshop assigned prerequisites to are shown in the table. Sub-goals with no operations listed is not included in the table. Additionally, some of the operations were not assigned prerequisites. These are also not included.

Assignment	Sub-goal	Operation(s)	Predicted Prerequisite
Attach (1)	Oriente Pen and Device (1.4)	Detect Groves and Knobs (1.41) Understand to align (1.42) Align Groves and Knobs (1.43)	Super User of Pen
Inject (2)	Attach Needle (2.2)	Lift Needle (2.21) Peel off paper (2.22) Lift pen with Device attached (2.23) Screw Needle onto Pen (2.24) Take of outer Needle Cap (2.25)	Super user of Pen Dexterity
Inject (2)	Set Dose (2.3)	Decide Dose (2.31) Dial dose (2.32)	Eyesight
Inject (2)	Inject (2.4)	N/A	Air shots
Inject (2)	Detach Needle (2.5)	N/A	Attitude towards needle re-use

Table 7.3: A table showing the predicted prerequisites and their affiliated operations, sub-goals and goals of the task. A parentheses () refers to the tag of the post it, which enables back-tracking.

Interpretation of Results from Task Analysis

The development and results of the task analysis seemed to suffer from the limited time and the extensive tasks. The participants did what they were told, but it also felt important in the situation that everybody understood what happened and why. Thus discussions about task analysis were allowed. As the facilitator, it was pointed out where they should focus to have some results applicable to the next assignment. This was achieved to some degree, but also left two whole tasks from the case incomplete. Additionally, none of the follow-up questions affiliated with the individual task were analyzed nor discussed.

Doing a task analysis does not seem fitting for the workshop format when the time is limited. Therefore, as suggested earlier, the tasks to analyze should have been less extensive. One could also argue that the familiarization could have been dropped in favor of more time for the actual assignment. Inspecting the results reveals that the participants understood the method and when asked, they found the added step interesting. The interest originated from the almost non-existent use of task analyses in their everyday work. They do make them sometimes, but they are not found as strong competitors for arguing in design or evaluation. The reasons for not using them is unknown. The added step was commented positively, as it created some potential for strengthening the use of task analysis in argumentation.

The discussion required for everybody to have a collective understanding of task analysis could imply different things. It implies that everybody has to have the same expertise within the domain of task analysis. When reviewing the literature, there are multiple ways and methods for task analysis, which demands that everybody participating in the development has the

same method in mind. Robust definitions of what to note down at which step, what to prioritize in some instances and such should be able to be identified through the description of the suggested method. While the task analysis is only a mean to reach a result useful for targeting other populations to sample from, it still seems that subjective and opinionated experiences have to be decreased.

If these minor issues are resolved the task analysis with the added step still seems like a very appropriate mean for obtaining the overall goal.

7.4 Predicting Human Errors with SHERPA

At this point in the workshop, some of the participants had to leave as they had to prioritize their work. This resulted in only 3 participants remaining. To make sure, that the participants got some insight into what they were doing, they were asked to only focus on one of the tasks they had completed. They continued with the task '*Attach*' and focused on the sub-goal '*Open Box*'. The ad hoc change was done, to simplify the assignment and allow for discussion. The argument for allowing discussion was that the internal discussion provided feedback on the use of the method. This type of data might have been lost if it was chosen to stress them and ask them to analyze the whole task.

Through discussion three different aspects surfaced. At first, it was questioned what would happen if an operation could be classified as more than just one type (action/retrieval)? Further how to access a probability of an error occurring, if there are no previous data? Lastly, it was discussed that a predicted error potentially could have multiple consequences and how this should be approached. The participants found that the ideal approach would be to note everything, including the multiple task classifications and consequences. It enabled an assessment later on, where it could be found what classification and consequence that would influence the most. Additionally, how the most influencing consequence could be avoided. For the question regarding the assessment of probability, there is no right answer. Preferably, other external data would available, but if not, the probability has to be subjectively assessed. While the predefined taxonomies were questioned, they were still used as intended, which validates that the use of SHERPA indeed is generic and applicable in multiple situations. Furthermore, the participants used their knowledge to provide additional information and extend the task classification. They used what they referred to as '*PCA*'. It covers perception, cognition, and action, and for each of them, they put a task classification. This extended version of the step could cover each task in higher detail, than the original step.

Results from SHERPA

As the participants only proceeded with the '*Attach*' task and the number of participants were reduced to three, the assignment was not solved fully. Instead, they focused on using the method for one of the sub-goals. As thereof, the SHERPA has through the workshop only been applied to the sub-goal '*Open Box*'. Table 7.4 shows what the participants developed. The questions that the participants initially asked was resolved by themselves. They, however, did not know how to assess the probability of the error. If absolutely no data is available, the best possible guess must be sufficient, as there is no other way of doing it. As the participants worked on the assignment, they realized that many of the consequences they had listed could be solved in several ways.

Task	Error Mode	Error Description	Consequence	Recovery	P	C	Remedial Strategy	Sample Consequence	Sample Criticality
1.1 Open Box	P: Retrieval (R1)	User do not see "open here"	Opens the box in an unintended way	Immediate	?		Make text larger. Move "Open Here" to the top of box	None	None
	C: Selection (S2)	Users open in an unintended way	Box and content is damaged. Device might be dropped	None	?	!	No box. Participant receives box opened	None	None
	A: Action (A9)	User struggles to open, as it requires too high forces	Frustration and abandonment of product	None	?	!	Participant receives box opened	None	None

Table 7.4: The results obtained by the participants in the development of SHERPA. P (Probability) is questioned, as they did not know. C (Criticality) is denoted by an exclamation mark but is originally judged on an ordinal scale (low, medium, high). The assessment of Sample consequences and Sample Criticality is based on a comparison between the previously predicted prerequisites and the content of this table.

The errors described for opening the box seems to be solvable by any of the three remedial strategies suggested. This connection between errors, consequences, and solution surprised, but also made other questions surface, e.g., How is the best possible remedial strategy chosen? It became apparent that the complexity increased when increasing the detail of the input, such as multiple classifications and consequences.

Interpretation of Results

In general, the use of SHERPA was fully understood by the participants and worked as intended. They did, however, pinpoint certain pitfalls, where definitions and guidelines are missing, which could imply the need for more thorough definitions. They also resolved many of the problems they experienced themselves, which proves that it is possible to bypass these pitfalls. It does require expertise and experience with such types of analysis, but it is not prohibitive to overcome these obstacles. Besides the very few pitfalls, the use of SHERPA was seamless and easy for the participants to understand. They did not get stuck in any of the steps but only suggested further optimization of some of the steps.

The solutions provided by the participants seems better than the original method as it accounts for even more aspects. However, it is also a resource intensive addition. The addition of PCA should only be chosen if the detail of the analysis is of high importance or if resources for the development of the analysis are at a sufficient level. The task analysis was deemed an appropriate mean for obtaining the overall goal, and so is SHERPA. The participants understood how the analyses of population and the task analysis functioned as means for using SHERPA. They also saw how SHERPA could be used for identifying other populations to sample from, and how a substitute sample provides a valid result and how the substitute can be defined as representative for the target population.

7.5 Complete Use of The Method

The workshop went as intended and provided input for optimization. The intended use is valid for both the workshop and the method. The participants did unfortunately never finish the task analysis nor the SHERPA. It seems unfair to comment on possible optimization, without having used the method thoroughly. Therefore, the task analysis and SHERPA are fully developed by the author. While this is no comparison to the workshop and the input from the participants, there might be unexplored issues or challenges that have not been addressed yet. It could also be, that there are only positive things to be said after both methods have been applied fully. The development is based on the initial work made by the participants. Thus the same case will be used. As the author's knowledge of people with diabetes is inferior to the people participating in the workshop, the analysis of the population is not performed.

A Fully Completed Task Analysis

To fully complete the task analysis, the work done by the participants have been reviewed. Also, the fourth task has been analyzed and put into the same structure as the one used by the participants. The third task is identical to the second task, and therefore, there was no need for replicating it. Instead, the questions related to the third task is described in coherence second task.

First Task - Attaching the Device

The first task was almost fully developed, and the only addition made is a sub-goal for the check of secure attachment. This is shown in figure 7.4. In the figure, there is no appearance of the predicted prerequisites, as these will be presented together with the rest of the tasks at the end of this section. The operation is added, due to the sub-goal not being an exact operation. By not having operations, the later use of SHERPA can be difficult as the taxonomies explicitly affiliate with the operations. The depiction of the added operations is originally from figure 7.2, which is a product of the participants' analysis in the workshop.

In addition to the actual task, the study used as a case involved three follow-up questions. All questions referred to the handling of the device and pen; How easy it was, if it could accidentally come off and how the participants figured out how to attach the two objects. None of these are deemed to be influenced by knowledge and experience.

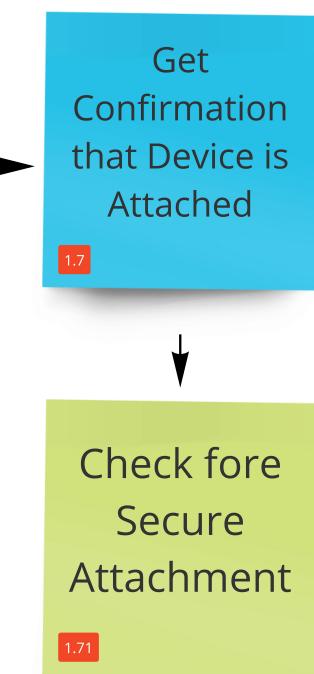


Figure 7.4: The added content for the Task analysis of the task "Attach". The sub-goal have been added a operational step.

Second Task - Injecting

For the second task, the participants missed out some of the operations required regarding actually performing an injection. Further, there were no operational steps for the sub-goal of detaching the needle. These have been added, as shown in Figure 7.5. The fact that the participants of the workshop did not add any of the operations for injecting is incomprehensible. Asking someone to inject themselves requires know-how and experience of doing it. The same applies to get off the needle and disposing of it correctly. Therefore, these have been added.

In the case-study, the person is told to imagine a scenario where he has downloaded an app. The information given is not deemed to have any influence on a potential substitute sample. For the follow-up questions, these only relate to how it was to inject. There could be some data on this question that can only be obtained by using a sample of people with diabetes. This is due to people with diabetes being able to compare everyday-injections with the injection performed in the task.

Questions for the third task

The third task asks the person to inject a smaller dose. The task itself is the same, but the wireframe app is going to show an inaccurate dose. After this, the person is asked about how he reacts and what he thinks about the inaccurate number. Such questions will not be answerable by a substitute sample, as it requires inside knowledge of insulin-doses. In addition to this requirement of knowledge, it can be very subjective whether it means something for the person or not. Some people with diabetes take 200 units, and some only take 3-4 units. It could be imagined that people usually taking 200 units would not care if it showed ± 1 unit, while patients on lower doses would react stronger to such inaccuracy.

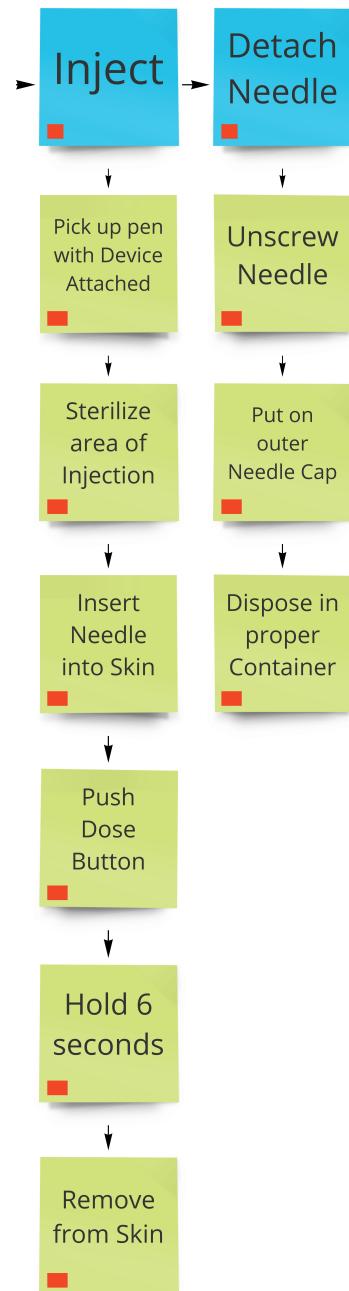


Figure 7.5: The additions made for the task analysis of the task 'Inject'. Two sub-goals have been reviewed and nine operations have been added. The sub-goals before are still the same, as the participant developed through the workshop.

Fourth Task - Detaching the Device

The participants never got to the fourth and final task. Therefore, this analysis is solely based on the author's subjective opinion and knowledge. The depiction of the analysis can be seen in Figure 7.6. It divides the task of detaching the device into four smaller sub-goals. In this task, the person has to attach the device to a new pen. This action seemingly requires the same operations as the first task. Therefore, no operations have been written for this sub-goal. Detaching the device requires the opposite of attaching it.

After having detached the device, the participant is asked how it was. Finally, they are asked to answer three questions, that they also answered at the beginning of the study. Two of the questions are answerable by anyone, as they refer to the use and knowledge of the device the person has just used. The last of the three is a judgment of whether they would like to use the device on a daily basis. This question is most likely not to be answered by any other than the target population.

Fourth Task

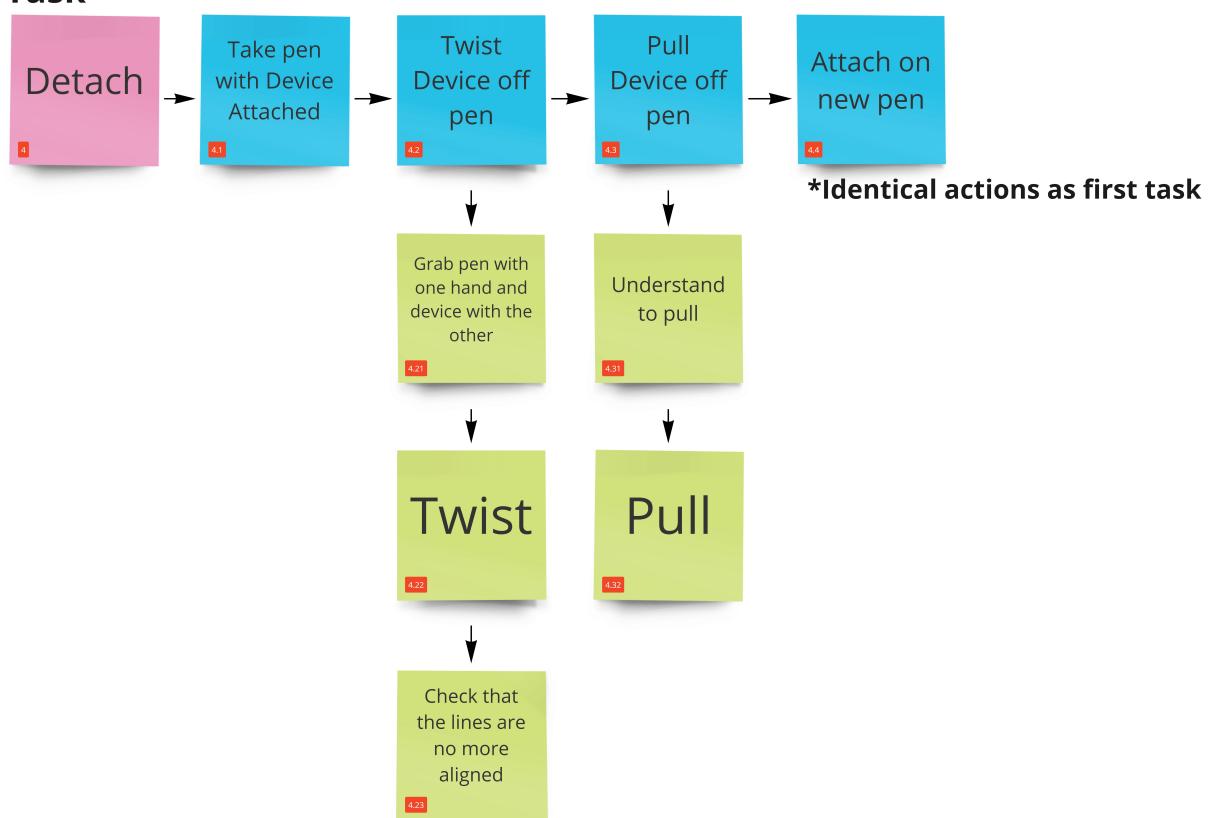


Figure 7.6: The fourth task of the case. Detaching the Device requires to reverse interaction as attaching it.

Predicted Prerequisites for All Tasks

The predicted prerequisites have not been included in the depiction of the task analysis. Instead, this final step summarizes everything through Table 7.5. It states the assignment, the sub-goals, the operations and lastly the predicted prerequisites. Some of the content is identical to Table 7.3 shown earlier. The similar content is due to the re-use of the results from the analysis the participants of the workshop developed. While the participant of the workshop did assign some required prerequisites, not all operations were analyzed. In this table, it ought to present all 'missing' prerequisites. Reading this table is done by looking at the left-most entry and read one row at the time. The predicted prerequisites are not bound to one single operation, but instead describes those needed for completing the sub-goal.

Table 7.5: The final result of the task analysis. It includes everything developed through the analysis and states the predicted prerequisites a person would need for completing the individual task, its sub-goals and the associated operations.

Assignment	Sub-goal	Operational Demand(s)	Predicted Prerequisite
Attach (1)	Open Box (1.1)	See "Open Here" (1.11) Understand where to lift (1.12) Lift Tab (1.13)	Normal eyesight
Attach (1)	Take out Device (1.2)	Recognize gap in pack to put finger (2.21) Move hand to gap and Device (1.22) Lift Device (1.23)	None
Attach (1)	Lift pen (1.3)	None	None
Attach (1)	Orientate Pen and Device (1.4)	Detect Groves and Knobs (1.41) Understand to align (1.42) Align Groves and Knobs (1.43)	<i>Super User of Pen</i>
Attach (1)	Push Device onto Pen (1.5)	Shape alignment (1.51) Understand to push (1.52) Push (1.53)	None
Attach (1)	Twist Device onto Pen (1.6)	See white lines (1.61) Understand lines should align (1.62) Twist (1.63)	None
Attach (1)	Get Confirmation that Device is Attached (1.7)	Check for secure attachment (1.71)	None
Inject (2)	Take off Pen Cap (2.1)	Recognize Cap part of Pen (2.11)	Pen knowledge

Continued on next page

Continued from previous page

		Understand the Principle of Removing Cap (2.12) Remove Cap (Pull) (2.13)	
Inject (2)	Attach Needle (2.2)	Lift Needle (2.21) Peel off paper (2.22) Lift pen with Device attached (2.23) Screw Needle onto Pen (2.24) Take of outer Needle Cap (2.25)	Super user of Pen Needle experience Needle handling
Inject (2)	Set Dose (2.3)	Decide Dose (2.31) Dial dose (2.32)	Eyesight Dose understanding
Inject (2)	Inject (2.4)	Pick up Pen with Device Attached (2.41) Sterilize area of injection (2.42) Insert needle into skin (2.43) Push Dose Button (2.44) Hold 6 seconds (2.45) Remove From skin (2.46)	<i>Air shots</i> Habit of injecting Hygiene knowledge
Inject (2)	Detach Needle (2.5)	Unscrew Needle (2.51) Put on outer Need Cap (2.52) Dispose in Proper Container (2.53)	<i>Attitude towards needle re-use</i> Needle experience Knowledge of disposal
Detach (4)	Take Pen with Device Attached (4.1)	None	None
Detach (4)	Twist Device off Pen (4.2)	Grab pen with one hand and device with the other (4.21) Twist (4.22) Check that lines are no longer aligned (4.23)	None
Detach (4)	Pull Device off pen (4.31)	Understand to pull Pull	None

While most of the prerequisites stated in the above table are understandable, some of them require some further explanation. These prerequisites are; 1) Super User of Pen, 2) Air shots and, 3) Attitude towards needle re-use. They are also marked in italic in the table.

Super User of Pen is taken from the participants from the workshop. It relates to the general knowledge of a person with diabetes. It can be hard to read from the table and even harder to understand out of context. Therefore, if super users are used as a prerequisite, it refers to knowledge and habit. In this case, the superuser prerequisite covers, i.e., which end of the device that is used for injecting or how to attach a needle to the pen.

Air shots are a typical action performed by some people with diabetes. It is a shot of insulin into free air. They do it before injecting, to check if the needle is clogged or anything else in pen is malfunctioning. This action is most likely not to be done by any other sample than people with diabetes unless the sample is also experienced with injectable medicine.

The attitude towards needle re-use is a very subjective matter for people with diabetes. It is for all patients stressed that re-using a needle is injurious to health. Many patients do however re-use needles as they find them expensive or troubling to change after each injection. Their attitude, in this case, will most likely influence their actions for the task, which most likely cannot be accounted for sampling-wise. Since the task analysis is fully developed, it is possible to use it as input for SHERPA.

A Fully Completed SHERPA

In the workshop, only a bare minimum of results was obtained through the analysis of SHERPA. The participants were productive and did what they were told, but the assignment was too extensive. To make up for this, a full SHERPA has been developed. All results are shown in Table 7.6. The table contains each sub-goal, which have then been classified. The system provided by the participants have been applied, and therefore each sub-goal have been classified in three different contexts, namely perceptual, cognitive and actions. Notice that '*actions*' also can be assigned other taxonomies than actions, e.g., a check or selection. Each classification has also been assigned an error mode taxonomy which is referred to by (X#). Further is a description of the error and the consequence thereof. Recovery from error, the probability of error and criticality of error is then noted. Lastly is the remedial strategy, followed by the consequences and criticality of the sample. For the SHERPA to make sense, each operational requirement for the sub-goal has been accounted for through analysis. In the table, the remedial strategy covers solutions to the overall problems of the sub-goal and not the individual operation.

Table 7.6: All results from the authors use of the methods. The results have been condensed in order to give an overview. From the left-most column is the given sub-goal. Each sub-goals have been assigned a perceptual, cognitive and handling error mode, which is then further broken down. In the last two columns, the consequences for using another sample and the criticality thereof, is described.

Task	Error Mode	Error Description	Consequence	Recovery	P	C	Remedial Strategy	Sample Consequence	Sample Criticality
1.1 Open Box	P: Retrieval (R1)	User do not see "open here"	Opens the box in an unintended way	Immediate	?		Make text larger. Move "Open Here" to the top of box	None	None
	C: Selection (S2)	Users open in an unintended way	Box and content is damaged. Device might be dropped	None	?	!	No box. Participant receives box opened	None	None
	A: Action (A9)	User struggles to open, as it requires too high forces	Frustration and abandonment of product	None	?	!	Participant receives box opened	None	None
1.2 Take out Device	P: Retrieval (R1)	Users do not use the gap in the packaging	The user will take the device out of the package in an unintended way	Immediate	M		Signify intended interaction	None	None
	C: None	N/A	N/A	N/A	N/A			N/A	None
	A: Action (A4)	Users struggle to get device out	The packaging might be damage	None	L		Give device without packaging	None	None
1.3 Lift Pen	P: None	N/A	N/A	N/A	N/A/N/A/N/A			N/A	N/A
	C: None	N/A	N/A	N/A	N/A/N/A/N/A			N/A	N/A
	A: Action (A6, A7, A10)	User might not believe device and pen are to be connected	User will have the wrong mental model	Immediate	L		If users are confused, provide a little more information	None	None
1.4 Orientate Pen and Device	P: Retrieval (R1)	Users doesn't see the knobs and groves	The user will struggle to get the device onto the pen		M	!	Shape device and pen for it to be possible to pick any position.	None	None
	C: None	N/A	N/A	N/A	N/A/N/A/N/A			N/A	N/A
	A: Action (A6, A7)	Users might not detect where the device should be attached.	Users will not be able to do 1.5, as the device is not in place. If they force it onto the pen, they might break either part.	Task 1.5	L	!	Ask participant if he thinks this is the place to attach the device. If they are forcing it, ask if they think a device would require such high efforts.	Substitute will most likely spend more time exploring the device	L

Continued on next page

Continued from previous page

1.5 Push Device onto Pen	P: Check (C1)	Users might not recognize the need for alignment	Users won't be able to push it all the way down, if it's not aligned	Task 1.6	L !	Better information in IFU or highlight the need for alignment on device	None	None
	C: Retrieval (R1)	The users misses the signifiers indicating the need for alignment	Users won't position the device as intended	Task 1.6	L !	Highlight signifiers	None	None
	A: Action (A4, A5, A8)	Users might just place it on top of the pen, and not get it in the right position. If they miss the alignment, they might try to force it.	Device won't be properly attached	1.6	L !	Prime participant to reconsider his actions	None	None
1.6 Twist Device onto Pen	P: Retrieval (R1)	Users won't see the white line on pen and device that needs to be aligned next	User won't figure out the need for twisting	Immediate	L !	Signify the need for a twist	None	None
	C: Selection (S2)	User will try other alternatives to twisting, e.g., pushing buttons	The device will not be attached properly. If they dialed a dose by mistake, pressure might rise in the cartridge with test-medium	Immediate	L !	Make sure participants does not try to use the push button with a dose dialed.	None	None
	A: Action (A3, A4, A8)	Users might twist in the wrong direction, overdo the twist or even omit to twist.	Device not properly attached. The wrong direction might break the device and/or pen	None	L !	Prompt the participants to verify that the device is properly attached. Stop participant if the forces used for attachment are too high.	None	None
1.7 Get confirmation that Device is Attached	P: Retrieval (R3)	User might only recognize that the device sits on top of the pen, and doesn't see that the white lines on pen and device do not align.	Participant will think it is properly attached, while it is not	Immediate	L	Prompting the participant to check again. If a twist has not been done, it will fall right off.	None	None
	C: None	N/A	N/A	N/A	N/A/N/A/N/A	N/A	N/A	N/A
	A: Check (C1, C2)	User will skip the check and firmly believe it is attached	Device might not be properly attached	Immediate	L	Prompting the participant to check again. If a twist haven't been done, it will fall right off.	None	None

Continued on next page

Continued from previous page

2.1 Take off Pen Cap	P: Retrieval (R3)	User won't retrieve what part of the pen that is a cap	User will try to pull/twist the dial	Immediate	L	Let participant explore device and find that only the cap can be removed from the pen.	Diabetics with pen-experience will most likely instantly remove cap	None
	C: Selection (S2)	Same as Retrieval	Same as Retrieval		L	Same as Retrieval	Same as Retrieval	None
	A: Action (A6)	Same as Retrieval	Same as Retrieval		L	Same as Retrieval	Same as Retrieval	None
2.2 Attach Needle	P: Retrieval (R1, R2, R3)	Users might not know how to attach the needle, if they are not experienced	Needle not being attached	None	L	Help attaching needle, as it isn't vital for the study	Pen-experienced users won't make any of these mistake	Pen-experienced sample or additional information before the task.
	C: Action	Inexperienced users will most likely need additional time for exploration	More time used for needle attachment	Immediate	L	Let participants have more time	None	None
	A: Action (A7, A9)	Failing to remove paper and screwing on the needle	Wrong needle attachment. The user might push the needle and then screw, leaving the paper to potentially clog	None	L	Ensure that paper is peeled off before attaching needle. Prompt them to attach a new needle and remove paper	Pen-experienced users will not make this mistake. Some people with diabetes will have 'creative' ways of doing it, e.g., holding the needle with the mouth.	Pen-experienced sample or additional information before the task.
2.3 Set Dose	P: Retrieval (R1)	Users might not see dose window	Users will not be able to set the dose to the number asked for	None	L	Let users explore. Potential help set dose, if not important for evaluation.	Additional information for participants or pen-experience.	None if dosing is not necessary to collect data on. Else pen-experienced users
	C: Action (A4)	User won't be able to validate correct dose and might dial too little/much	Wrong dose	None	L	Pre-set dose	Additional information for participants or pen-experience.	None if dosing isn't necessary to collect data on. Else pen-experienced users
	A: Action (A4, A6, A7)	Fail to dial in the right direction or dialing the right part of the device.	Wrong dose	None	L	Pre-set dose	Additional information for participants or pen-experience.	None if dosing isn't necessary to collect data on. Else pen-experienced users
2.4 Inject	P: Retrieval	Users won't know to sterilize and hold in device	Unhygienic injection and full dose not delivered	None	L	Give additional information	A substitute sample will need additional information and most likely help	A sample familiar with injections to be preferred
	C: Selection (S2)	Inexperienced users will most likely hesitate to inject	Users will take longer to inject	Immediate	L	Allow additional time for un-experience users	A substitute sample will need additional information and most likely help	A sample familiar with injections to be preferred

Continued on next page

Continued from previous page

	A: Action (A4, A8, A9)	Users fail to hold dose-button down	Not getting full dose	Immediate	L	Additional information	A substitute sample will need additional information and most likely help	A sample familiar with injections to be preferred
2.5 Detach Needle:	P: None	N/A	N/A	N/A	N/A/N/A/N/A		N/A	N/A
	C: Selection (S1, S2)	Without further information, failure putting back on the outer needle cap and disposing of the needle in a fashionable manner happens. Users reusing the needle will not do any of the two operations.	Wrong disposal of needle or user not disposing the needle at all. Contamination of needle and possible danger for others	None	L !	Provide information of proper disposal.	Inexperienced users might need additional information	None unless detaching the needle is important for the study
	A: Action (A3)	User might stab himself with the needle, if not putting on outer needle cap	Failure to safely detach the needle	None	L !	Prompt participants to do it safely and use outer needle cap	Inexperienced users might need additional information	None, unless detaching the needle is important for the study
4.1 Take Pen with Device	P: None	N/A	N/A	N/A	N/A/N/A/N/A		N/A	N/A
	C: None	N/A	N/A	N/A	N/A/N/A/N/A		N/A	N/A
	A: None	N/A	N/A	N/A	N/A/N/A/N/A		N/A	N/A
4.2 Twist Device off Pen	P: Check (C1)	User fail to check that the white lines on pen and device are still aligned	Won't be able to perform 4.3	4.3	L	Let user explore the device and pen.	None	None
	C: None	N/A	N/A	N/A	N/A/N/A/N/A		N/A	N/A
	A: Action (A3, A6, A7)	User fails to twist in the right direction or accidentally dials a dose when trying to twist.	The device won't be detached. Too high force used twisting the wrong direction might break it. Also a dose might be set	Immediate	L	Ask if they think a device would require such high efforts to detach.	None	None
4.3 Pull Device off pen	P: None	N/A	N/A	N/A	N/A/N/A/N/A		N/A	N/A
	C: None	N/A	N/A	N/A	N/A/N/A/N/A		N/A	N/A

Continued on next page

Continued from previous page

A: Action (A4)

User fails to apply the correct amount of force when pulling

The device stays attached or flies off

Immediate L

Let user explore if force used is too low. If it's too much they most likely will be surprised

None

None

Reviewing the Author's Use of The Method

Using both the task analysis and SHERPA single-handed was surprisingly easy. There are no downsides to doing it alone, other than not being able to discuss what goes into the analysis. The omission of discussion makes the results much more subjective but is inevitable when only one researcher conducts the analysis. Even though it was easy to apply the methods some parts of the methods worked more optimal than others. The task analysis was straightforward and did not cause any difficulties.

Many of the steps in SHERPA was not causing any difficulties either, but some issues became apparent. The use of PCA, suggested in the workshop, was very helpful and forced one to think about other aspects. It was however tough to apply the taxonomies for the perceptual- and cognitive levels. On the perceptual level, almost everything had to be put into the retrieval error mode. A need for a larger taxonomy for related perceptual errors was present. While the application of taxonomy was a bit troublesome the issue was remedied through the description of consequences. When allowed to describe the actual problem and consequence, the issue of assigning a taxonomy became negligible. The described consequences only remedied the issues of the perceptual aspect.

The biggest problem was applying the taxonomies to the cognitive aspect. The taxonomy is not suited for including experience, lifestyles or habits that might influence a participant's actions. Using a substitute sample entails that the participants have different mental models and thereby will experience the product differently. They do not have the same knowledge or experience, which causes their view on the product, to differentiate from people from the target population. Many of the problems one imagine to experience if not using a sample from the target population is that the mental models the participants have and develops are different. Additionally, the consequences of the cognitive aspect most of the time lead to the substitutes requiring more time for product exploration. It can be hard to assess this as errors. Instead, these highlights some of the difficulties that might not lead to an error but instead prolongs the exploration and learning of the device in scope.

Finally, SHERPA seems to take almost everything into account. The only missing aspect is the absence of errors not related to the task itself, e.g., participants dropping the device on the floor by mistake. The case used in the workshop was conducted by the author in collaboration with other employees from Novo Nordisk last year. While many of the mistakes, that were observed in this study are identifiable through the method, the method cannot be used for unpredictable errors. Last year, a participant somehow managed to 'shoot' the device off the pen, meaning that the device detached with high enough force to fly off the pen. The method is still deemed applicable as a mean for identifying substitute samples. The method will never be able to predict unpredictable and infrequent errors, but neither will any other methods

without user inclusion.

Using SHERPA for identifying the opportunity for substitute sampling and the criteria thereof is deemed possible. The method does lack some aspects, and the taxonomy could be further developed with the scope of changing the area of use to sampling. It is, however, in its present state, still useful.

7.6 Summary of Findings and Suggested Changes

Through the workshop and using the methods myself, a greater understanding has been obtained. The workshop allowed for insights of others using it. It showed that domain expertise has great influence on the approach the person has to the methods. Using the methods single-handed allowed for insights that were not possible to gain through the workshop. The experience and insights obtained are summarized in this section.

Using personas

Personas can be very valuable to have, as they can be drawn upon even though the user evaluation is not included. Developing personas is, unfortunately, a very extensive and resource intensive method. It requires efforts that most likely will not be possible to prioritize in small business' with strict deadlines and other assignments with high priority.

An alternative to personas is identity models (Holtzblatt and Beyer, 2014). Their purpose is to structure a model for the people being designing for. Identity models are much more situational than personas, which implies that the use of them also are bound to the single evaluation. Personas seem more generic and have a more diverse use. As with personas, they are based on knowledge of the population and interpretations of observations and articulations previous gathered from other studies. From this knowledge and data patterns are identified, which are used in the development where it is tried to encapsulate the essence of the target population (Holtzblatt and Beyer, 2014).

A full identity model covers three overall aspects which are; '*I am*', '*I Plan*' and, '*I Like*'. Using identity models as a mean for finding out what kind of sample to use, does seem applicable when resources do not allow for personas. Furthermore, if the identity model is deemed too vague and contains too little detail, it is possible to connect it to other models. These other models try to uncover other aspects that have an influence. They particularly try to uncover relationships, collaboration, and sequences. Possibly, connecting the identity model to the other models will require more resources and time, which has to be taken into consideration; is it worth connecting several models over making personas?

The suggestion of using identity models over personas, should not be perceived as a replacement of personas. Instead, it is a '*discount*' alternative when it is questionable whether enough resources can be allocated for developing personas. While identity models is a less extensive user model to apply, it has to be considered if the delimited method for personas can be to better use. This consideration should be done because it seems that identity models require more resources than the delimited persona.

Using Task Analysis

Using the task analysis as a mean for identifying substitutes for sampling worked as intended. The participants from the workshop only had positive feedback for the addition of the last step. They found it to strengthen the use of the task analysis as a tool of argumentation. Additionally, it seemed straightforward for the participants to predict prerequisites for the individual operations. Regarding issues of applying the task analysis, two issues were identified. Some of the prerequisites that are needed for an operation can be so domain specific (e.g., air-shots) that the prerequisite needs additional explanation. While this is not a problem if everybody has the specific knowledge, it can be hard to obtain the full meaning of the task analysis, if only relying on the condensed table of results.

The last issue was that the participants of the workshop demanded the method to be less academic and theoretical. While using the method single-handed this did not occur to be a problem. The easier use could be for several reasons but is most likely due to the method not being put into a workshop format. Using the method out of the workshop seemed very manageable. Additionally, there were no others to discuss with, making it somewhat more efficient. While being more efficient when analyzing alone, it could be that the results suffer from the more subjective opinionated data. While agile methods often are preferred in business' it is also a trade-off. Making the task analysis less academic and theoretical will potentially diminish the effect it can have in an argument.

One last thing was observed. The different approaches to the task analysis were very dependent on the domain expertise the participant already had. It is not possible to include domain expertise into the optimization of the method, but being a team creating task analyses will most likely help to cover all perspectives. People not being familiar with the method will most likely not succeed using it, as it requires knowledge to make a good task analysis. This also implies having flair for usability and breaking down tasks into meaningful goals. The need for being familiar seems valid for everything, as beginners in a domain will make mistakes.

On the more positive side, all the participants who could be said to be experts in user experience and usability seemed to have very identical approaches to the task analysis. It suggests that it is possible to create a unanimous understanding of the method, which could indicate that subjective opinions do not affect the use of the method. This statement cannot be generalized, as it would require a lot larger sample but evidence that it might be true did appear through the workshop.

Using SHERPA

SHERPA enabled achieving the overall goal. From the results of all analyses, it was possible to identify when a substitute sample was possible. Further, it identified the consequences and

the criticality of using a substitute sample. From these consequences, a set of new sample criteria are developed. The new sample criteria should be broader and extend to others than the target population for it to be useful. The criteria define the new sampling framework and enable argumentation for being able to sample from other populations. The development of the new sampling criteria should relate to the consequences. If a consequence states that a substitute sample will not be able to perform the task as intended, due to X reason, the new sampling criteria for this particular sub-goal should include X. All consequences together should end up being the new criteria for the sample.

Regarding the use of SHERPA, using the task analysis as input for SHERPA was done with ease. Predicting the sample consequences was not tricky, and neither was the sample criticality. It was, however, harder to apply the predefined taxonomies. The taxonomies worked as intended in the original method, but applying the PCA model to the error mode prediction created some difficulties. These difficulties mainly revolved around a lack of fitting taxonomies for the cognitive aspects. Later it became apparent that accounting for emotional- and user experience aspects were troublesome. Neither the HTA or SHERPA is initially designed for analyzing the human aspects but instead focuses on the tasks, the required interaction and, the design-induced errors. The goal of the original methods excludes human-induced errors, which in fact are what the substitute sampling is about. Omitting the cognitive aspects in the analyses does not make it impossible to use and succeed. The results will still be valid but only concerns the tasks. It is imagined that predicting anything sampling-wise will be difficult if omitting the cognitive aspects.

General Suggestions and Overall Impression

In general, there are not many changes to the methods, as they proved to be valuable as they are. The methods gave the impression that it is possible to use substitute samples. Additional, it is possible to identify the new sampling criteria for this substitute sample. Furthermore, it showed that it was possible to identify the consequences of using a substitute sample through the analytically approaches. Opposite literature and previous studies, it seems that an alternative to sampling exists in-between the perfect representative sample and the typical convenient sampling. It is possible to adopt a sliding scale when referring to sampling and it would be entirely acceptable to envisage the flexibility of sampling.

Even though sampling has proven to be more flexible than the literature suggests, the need for a less academic and theoretical approach was still wished from the participants of the workshop. The method was deemed more comfortable to use when doing it single-handed, which could suggest that the wish from the participants were due to the method being put into a workshop. Putting it in the workshop made it very step-by-step alike.

To make the methods agiler and less extensive the 'discount' personas could be used. For the

HTA and SHERPA, only suggestions for optimization can be provided. There are not enough resources for validating if the suggestions work.

If the sole purpose is to identify when, how and why to use substitute samples some of the steps in both methods could be reviewed. It could even be that some of the steps could be less extensive or even removed. Reducing the extent or removing steps is however only a valid suggestion if there is no wish to keep the additional data the methods provide.

In the HTA plans could possibly be removed entirely, as the impact of the sequence of intended handling is deemed situational. Removing the plan analysis does have to be considered for each evaluation as it could also have an enormous impact. All other steps seem to be necessary because the detail of the task analysis transfers directly into SHERPA.

Concerning SHERPA the only step deemed unnecessary in the use of substitute identification is the Recovery Analysis and the affiliated Remedial Strategy. It is not a need-to-have but instead a nice-to-have element. Arguably the assessment of probability and criticality of the errors could be removed, as well because they do not have a direct connection to the identification of substitute samples. Yet, both of these have an indirect influence. If the probability and criticality of an error are deemed high when using the target population, this will most certainly appear in the consequences of a substitute sample. Also, if an error is assessed low on these two scales, then it shows that this error most likely will not happen for the target population. If a substitute sample is then used and the error happens frequently, it has to found why because the substitute sample clearly is not like the target population for this specific sub-goal.

While agility also can cover aspects like the sequence of steps and so on, it is thought that the order of the steps most likely will not benefit from being changed. Furthermore, the order seems entirely cemented as the steps are linked together and sometimes forced to be done in sequence.

Definitively it has to be stated that there is no right or wrong answer to how optimization would be done, without further testing and validation. Further testing is not an option in this project, and therefore the optimization of the method can only be based on speculations. The workshop, the single-handed use and everything leading up to this point, however, enables a discussion on the subject of sampling and substitute sampling. Therefore, the next chapter is dedicated to discussion.

Chapter 8

Discussion of Sampling

The method developed for identifying when and how to use substitute samples have been discussed and interpreted. While the method itself have been undergoing discussion and interpretation, the results have not been put into perspective. Putting the results into perspective is done by setting the results in contrast to the literature, previous studies and the articulations given upon the subject from the interviews.

8.1 Literature, Previous Studies, and The Results

Several points have been made throughout the thesis. Results showed how sampling possibly could be approached if there is no access to the target population or a convenient sampling method would not satisfy. As literature divides sampling into these two end-points, it can be difficult to understand why there are no alternatives, when it seems possible to use substitute samples. Methods as the convenient, the theoretical and the purposeful does cover different strategies for sampling, but they are all based on the presumption that the target population is somewhat accessible.

The purposeful and the theoretical sampling method could easily contain approaches to take when the target population is not available. A method like the one developed could be used for establishing more structure in the theoretical and purposeful sampling methods. It shows the purpose of the evaluation of each task. It shows how and why the purpose is to be fulfilled by another sample and it shows the prerequisites that affect the participants' answers. It functions as a tool for obtaining the sampling framework that is needed for fulfilling all purposes an evaluation has.

While both the theoretical and purposeful sampling method mentions that the goal is to sample with respect to purpose, they do not provide the means for doing so. While personas are a great tool for understanding users and including the user needs into an establishment of sampling criteria, there is no mentioning of these or any other tools to be used. It is strange that the way of sampling and recruiting participants for evaluation happens on a sliding scale, but in literature, sampling is divided into representative or convenient.

The literature on sampling becomes even more incomprehensible when reviewing previous academic work and talking with people working in corporations. In situations where the burden of proof is very high, not many alternatives to using the target population seem ade-

quate. It is not that the results would not be able to be obtained by a substitute, but merely the doubt that always will exist. Is the substitute representative for the target population? There is no real way of telling unless the evaluation is conducted on both a substitute and target population. A comparison will be needed every time, for it to be verified entirely that the substitute is alike. Additionally, such comparison should be made for each evaluation if one has to be certain. This undermines the idea of spending resources to identify a substitute. The procedure of identifying a substitute is really to expand the sample framework and criteria by arguments obtained through robust and rigorous analyses, which there is no mentioning of in literature nor previous studies. Previous studies using alternatives to the target population seems to rely on the convenience aspects mostly and typically proceeds with what's available resource- and time wise.

8.2 The Method and Further Verification

While the use of the method for identifying substitute samples have been deemed useful, there are still unanswered questions and issues that have not been accounted for. The issues that have not been accounted for is the optimization of the method. There is a definite need for the methods to be less academic and theoretical if they have to be implemented in everyday work routines. It cannot be said how this should be done without further testing and trying out the suggested alternatives of the methods that potentially could ease up the use.

The academic and theoretical aspects of the method might be one thing, but the methods of identifying a substitute still have not been verified - only that the methods provide the intended output, and that the sample criteria from there seem promising. For the methods to succeed fully, a sample should be drawn with the help of the method, and a study should be conducted. It might prove other problems that have not been put a focus on yet. The same applies to the analysis of the data, and such evaluation would provide. Earlier, it was stated the data might have some different characteristics, due to evaluating without the target population. It might be, that it can be challenging to keep an objective point of view. It also may be difficult to separate the data concerning what is answerable and what is not, by the substitutes.

These speculations remain to be speculations until they have been tested. There is no clear answer without results, and there is undoubtedly no generalizable answers until then. While these speculations are incredibly valid, it should be stated, that the use of personas, task analysis, and SHERPA combined do result in higher knowledge of population and evaluation. This knowledge also contributes to developing and expanding the sample framework.

8.3 Related Issues

While the verification of the full use of the methods still misses, there are other aspects in which speculations have to be done. This especially regards sample size and the taxonomies of the methods.

Some speculations have been made about sampling size, and the rule of thumb was to follow the ordinary guidelines. There might be a need for slightly larger samples when sampling from another population, as other aspects might influence the data. The presumption that the ordinary size of a sample can be used is not known for sure, as there is no research conducted in the area. It is however suspected that the conventional assumptions of sample size are adequate to a certain degree. All actions for obtaining a normally distributed set of data or a completely saturated set of qualitative data should not be different, as these rules are used today. Furthermore, these rules are generic and have been applied to multiple types of evaluations in previous research and evaluations. There could, however, be higher variances in data, due to the sample being put together by multiple populations fitting the scope of the evaluation. This would be the argument for using larger samples, as higher variances implicate less unanimously opinions and answers from participants. The issue of sample sizes will most likely be solved through iterations of use of the method. It will be quite clear whether or not the data collected through an evaluation with a substitute sample is adequate and results in a complete set of data.

The issues regarding taxonomies might be a bit more unclear. Stating that the method for identifying substitute samples and the criteria should be obtained through a robust and rigorous method also entails a taxonomy that can cover everything. Having a taxonomy that does not account for all aspects, could potentially cause the users of the methods to fabricate new entries in the taxonomy, fitting their objectives. While this, of course, is positive for the individual case it might cause very high variances in the use of the method. This is due to each user of the method having a unique taxonomy. If this is the case, it could be that the results from the methods heavily rely on these. This ruins the robustness and possible influences the final results.

It could however also be that the method itself is so robust, that these individual based entries in the taxonomy do not skew the data. If this is the case, the agility of the method seems very promising, as it can be tailored to the individual's needs. Once again, this calls for further research. If such research were to be conducted, it would, however, seem wise to group the taxonomies and potentially develop a set of new generic taxonomies. This would prevail wrong use of the taxonomies. The reason for the taxonomy problem to exist is deemed to be caused by the original area of application. The taxonomy in its current state does not accommodate applying it, with the scope of identifying substitute samples. Further, it does not focus on the more human variables, such as experience, habits or feelings. The method is not to be blamed for lacking these inputs, as they originally were not thought to be included.

8.4 Maintaining Context and Motivations

When making user evaluations, it can be quite controversial to claim that context and motivation can be retained during the evaluation. It is found hard to believe that all motivation that drives the participants in ordinary days can be sustained in an evaluation. The motivation for taking insulin is quite clearly a motivation to keep being healthy and not to get any related diseases. A motivation this strong does not seem to be sustained throughout an evaluation. It is not that the participant loses this motivation, but merely that the motivation is not applicable to the test scenario. Asking a user to inject a certain dose with a prototype will not necessarily rely on the motivation of being healthy. Instead, the motivation for doing the injection could be politeness as the participants were kindly asked to do it.

While it arguably is not possible to sustain ordinary motivations in evaluations, there are of course some influence, that cannot be measured. The same might be the case for context. While context surely will have an impact, it is also deemed impossible to sustain everyday context when the participant uses the devices in the evaluation.

Motivations and context are possibly dictated by the type of evaluation itself, as most of the time the participants are asked to do X with Y, without any further consequences. The variables associated with having diabetes (or any other lifestyle changing disease) simply cannot be put into a test-scenario. Yes, it is possible to remedy some of the adverse effects a test setup will provide, but arguing that good evaluation design can completely uphold motivation and context are nonsense.

The reasoning for discussing the value and possibilities regarding motivation and context is that these characteristics and prerequisites of the population maybe can be disregarded in some way when trying to obtain a substitute sample. This, of course, will not be an option if developing personas, but with the only goal of identifying substitute samples, it might be possible. For it to be possible, it has to be investigated which kinds of motivation that influences. Next, if they, at all, can be sustained in an evaluation environment.

8.5 Future Work and Needs

While the previous parts of the discussion also state some potential future work, there are still aspects that have not been included at all. Through the interviews with Morten and Tina, it became apparent that sampling is not just knowledge of the population. It is the design of evaluation, it is context, and it is fidelity of prototypes. The complexity of sampling is enormous, and each time a given aspect is investigated new aspects and questions are found.

Whilst many of these questions are yet to be discovered and answered, the most predictable

one is the fidelity of the prototypes. In the design of the method, the evaluation is analyzed which implicitly involves the fidelity of the prototype. It can however not be answered if fidelity will be an issue when using substitute samples.

A presumption could be that having participants not associated with the target population and exposing them for a low-fidelity device specifically designed for the target population would cause issues. It can be challenging enough for participants to familiarize themselves with an entirely new product, they never thought they should experience. Now imagine a prototype of low fidelity. This only further complicates what one are asking the participant to relate to. These situations cannot be taken care of by the method. They can be somewhat avoided by proper design of scenario and use of cases, but the real influence is unknown.

The same process of thinking applies to the contextual influences. Can substitute sampling be used in both laboratory and field testing? Field testing is known for having a high contextual detail, but is it possible to maintain this, when the participants are not from the target population? When using the substitute sample, everything will be staged. The use of substitute samples in evaluation where context is essential requires a proper design of the evaluation, as context typically also is defined by the kind of users a product has.

In the end, these suggestions for further work is to be conducted, if the method is used for selecting a sample. Before this, the method itself can easily be tested and researched further. The room for improvement regarding agility and taxonomy should be prioritized before the method is used for substitute sampling. The method might work in its current state, and it might be possible to successfully obtain a substitute sample by using it, but is it also evident that improvements can be developed.

Chapter 9

Concluding on Substitute Sampling

The thesis ought to answer how sampling can be approached under challenging conditions. From literature it became apparent that sampling is perceived as two cemented categories, namely representative and convenient. It was however also apparent that sampling possibly could be perceived as a sliding scale. A lack of how to identify and use this sliding scale of sampling in evaluations was however not evident. Through reviews of previous studies and interviews with people from corporations, some aspects became apparent and helped in the understanding of sampling as a sliding scale.

When it is unwanted to expose a product outside the business framework due to confidentiality or secrecy, the conclusion is that it is a self-created problem. It is a subjective weighting of the value gained from not exposing it to the value of user inclusion. If it is imperative that the design is not exposed, it was suggested that features and functionality of the product, could be put into a new 'fake' design, allowing for user inclusion without exposing the design. Very importantly, this approach eliminates the use of evaluations with objectives related to an assessment of product appearance and product adoption. E.g., questions asking if the user could imagine buying the product and having it in their home.

The influence of qualitative and quantitative approaches to sampling generally didn't have an impact for using substitute samples. What seemed most important, was that the optimal sample size changes based on the type evaluation.

When it is not possible to draw a representative sample from the target population, several options remain. It has not been mentioned before, but it is an option to be satisfied with a convenient sample. The convenient sample is the easiest and least resource intensive sample but also results in reasonable doubt whether the results are generic for the target population. Another approach is to use the method designed through the thesis. By applying a rigorous analytic methodology to the issue, it is possible to expand the sampling framework. Identifying new criteria for the sample allows populations that share some of the characteristics of the target population to be included in a potential sample. The analytic approach includes the use of personas, as these increase the knowledge of the target population. It can be used in several phases of development besides expanding the sample framework, and its strength lies in the representation of the prerequisites, experiences, knowledge and needs the target population has.

Next in the analytic approach is the task analysis. It was found, that type and goal of evaluation profoundly influence the type of sample. Through task analysis, it is identified what the tasks demand from the user. The task analysis condenses the tasks into operations necessary for doing the task as intended. Furthermore, the task analysis map the needed prerequisites for each task. The task analysis is then used as input for SHERPA, which in short applies a methodology to the prediction of errors. The combination of personas, task analysis, and the predicted errors can be used for assessing the consequences sample-wise and assess the criticality of using a sample from the target population.

Through assessment, it is possible to find what kind of data a substitute sample can provide and what kind of data that is restricted and only obtainable by the use of the target population. Lastly, this is compared to the goals of the evaluation. The comparison further enables seeing how critical it would be not to use a sample from a target population for each operation. Altogether, the methods provide input for when a substitute sample can be beneficial. Further, it identifies a new set of sampling criteria. In the end, this allows for making informed decisions for the sample instead of relying on convenience. As a conclusion, it is found that literature does not show the full aspect of sampling. There is still room for improvement of the method designed through the thesis, and it still has to be fully verified. The method does, however, lead to a rigorous procedure, that shows how sampling can be perceived as a sliding scale, instead of two cemented options.

Even though the method works as intended, some problems have been experienced. The project work have been subject for some unforeseen issues, such as broken stimuli that dictated a new type of validation of the method. In general, the lack of literature on substitute sampling (or surrogate evaluations) have been troubling, as existing methods for hard-to-reach sampling are very vague. While issues have been present through the project, most aspects have been very positive and all people who have given input throughout the project have been nice to work with.

As a final conclusion, the subject of sampling seems very easy to approach, when the target population can be accessed. It is however very difficult if it is not. The method designed through the project are a strong competitor to the existing methods for hard-to-reach samples and seems very promising. It is hoped that more people will perceive sampling as a sliding scale and start questioning what can be done in difficult situations. This would hopefully increase the amount of methods and literature on the hard-to-reach sampling.

References

- Cooper et al., 2007.** Alan Cooper, Robert Reimann and David Cronin. *About face 3: the essentials of interaction design*. John Wiley & Sons, 2007.
- Coyne, 1997.** Imelda T. Coyne. *Sampling in qualitative research. Purposeful and theoretical sampling; merging or clear boundaries?* Journal of Advanced Nursing, 26(3), 623–630, 1997. ISSN 1365-2648. doi: 10.1046/j.1365-2648.1997.t01-25-00999.x. URL <http://dx.doi.org/10.1046/j.1365-2648.1997.t01-25-00999.x>.
- David Dayton, 2003.** David Dayton. Audiences involved, imagined, and invoked: trends in user-centered interactive information design. In *Professional Communication Conference, 2003. IPCC 2003. Proceedings. IEEE International*, pages 9–pp. IEEE, 2003.
- Dumas and Redish, 1999.** Joseph S Dumas and Janice Redish. *A practical guide to usability testing*. Intellect books, 1999.
- Field and Hole, 2003.** Andy Field and Graham Hole. *How to Design and Report Experiments*. SAGE publications Ltd, 2003.
- Holtzblatt and Beyer, 2014.** Karen Holtzblatt and Hugh Beyer. *Contextual design: evolved*. Synthesis Lectures on Human-Centered Informatics, 7(4), 1–91, 2014.
- Rasmus Jensen. Connecting Product Attributes to User Experiences, Development of methodology through a case with medical injection devices. Master's thesis, Aalborg University, 2017.
- Plinio Thomaz Aquino Junior and Lucia Vilela Leite Filgueiras, 2005.** Plinio Thomaz Aquino Junior and Lucia Vilela Leite Filgueiras. User modeling with personas. In *Proceedings of the 2005 Latin American conference on Human-computer interaction*, pages 277–282. ACM, 2005.
- Kawakita, 1975.** Jiro Kawakita. *The KJ method—a scientific approach to problem solving*, Technical report, Kawakita Research Institute, Tokyo, 1975.
- Kelly, 2010.** Brian C Kelly. *Sampling and recruitment issues in qualitative drug research: Reflections on the study of club drug users in metro New York*. Substance Use & Misuse, 45(5), 671–683, 2010.
- Kennan et al., 2012.** Danielle Kennan, Allyn Fives and John Canavan. *Accessing a hard to reach population: Reflections on research with young carers in Ireland*. Child & Family Social Work, 17(3), 275–283, 2012.

- Jesper Kjeldskov, Mikael B Skov, Benedikte S Als and Rune T Høegh, 2004.* Jesper Kjeldskov, Mikael B Skov, Benedikte S Als and Rune T Høegh. Is it worth the hassle? Exploring the added value of evaluating the usability of context-aware mobile systems in the field. In *International Conference on Mobile Human-Computer Interaction*, pages 61–73. Springer, 2004.
- Kjeldskov et al., 2005.** Jesper Kjeldskov, Connor Graham, Sonja Pedell, Frank Vetere, Steve Howard, Sandrine Balbo and Jessica Davies. *Evaluating the usability of a mobile guide: The influence of location, participants and resources*. Behaviour & Information Technology, 24(1), 51–65, 2005.
- Koerber and McMichael, 2008.** Amy Koerber and Lonie McMichael. *Qualitative sampling methods: A primer for technical communicators*. Journal of business and technical communication, 22(4), 454–473, 2008.
- Kothari, 2004.** Chakravanti Rajagopalachari Kothari. *Research methodology: Methods and techniques*. New Age International, 2004.
- Kvale, 2004.** Steinar Kvale. *InterView: En introduktion til det kvalitative forskningsinterview*. Hans Reitzels Forlag, 11 edition, 2004.
- Vincent Laagland. Patient Empowerment in Type 1 Diabetes Care. In the Early Stages After Diagnosis. Master's thesis, Delft University of Technology, 2017.
- Cynthia LeRouge and Jiao Ma, 2010.* Cynthia LeRouge and Jiao Ma. User profiles and personas in consumer health technologies. In *System Sciences (HICSS), 2010 43rd Hawaii International Conference on*, pages 1–10. IEEE, 2010.
- Lincoln and Guba, 1986.** Yvonna S Lincoln and Egon G Guba. *But is it rigorous? Trustworthiness and authenticity in naturalistic evaluation*. New directions for evaluation, 1986(30), 73–84, 1986.
- MacNealy, 1999.** M.S. MacNealy. *Strategies for Empirical Research in Writing*. Allyn and Bacon, 1999. ISBN 9780205272532.
- Marshall, 1996.** Martin N Marshall. *Sampling for qualitative research*. Family practice, 13(6), 522–526, 1996.
- IM Moczarny, MR De Villiers and JA Van Biljon, 2012.* IM Moczarny, MR De Villiers and JA Van Biljon. How can usability contribute to user experience?: a study in the domain of e-commerce. In *Proceedings of the South African Institute for Computer Scientists and Information Technologists Conference*, pages 216–225. ACM, 2012.

- Muller et al., 2001.** Michael Muller, David R. Millen and Carol Strohecker. *What Makes a Representative User Representative? A Participatory Poster*. Human factors in computing systems, 2001.
- Nicolucci et al., 2013.** Antonio Nicolucci, Katharina Kovacs Burns, Richard IG Holt, M Co-maschi, Norbert Hermanns, H Ishii, A Kokoszka, Frans Pouwer, SE Skovlund, H Stuckey et al. *Diabetes Attitudes, Wishes and Needs second study (DAWN2™): Cross-national benchmarking of diabetes-related psychosocial outcomes for people with diabetes*. Diabetic medicine, 30(7), 767–777, 2013.
- Jakob Nielsen and Rolf Molich, 1990.* Jakob Nielsen and Rolf Molich. Heuristic evaluation of user interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 249–256. ACM, 1990.
- Platt et al., 2006.** Lucy Platt, Martin Wall, Tim Rhodes, Ali Judd, Matthew Hickman, Lisa G Johnston, Adrian Renton, Natalia Bobrova and Anya Sarang. *Methods to recruit hard-to-reach groups: comparing two chain referral sampling methods of recruiting injecting drug users across nine studies in Russia and Estonia*. Journal of Urban Health, 83(1), 39, 2006.
- Rivers, 2008.** Douglas Rivers. *Understanding people: Sample matching*. Palo Alto: YouGov\Polimetrix, 2008.
- Rogers et al., 2011.** Yvonne Rogers, Helen Sharp and Jenny Preece. *Interaction design: beyond human-computer interaction*. John Wiley & Sons, 2011.
- Rubin and Chisnell, 2008.** Jeffrey Rubin and Dana Chisnell. *Handbook of usability testing: howto plan, design, and conduct effective tests*. John Wiley & Sons, 2008.
- Shaghaghi et al., 2011.** Abdolreza Shaghaghi, Raj S Bhopal and Aziz Sheikh. *Approaches to recruiting ‘hard-to-reach’populations into research: a review of the literature*. Health promotion perspectives, 1(2), 86, 2011.
- Silverman, 2015.** David Silverman. *Interpreting qualitative data*. Sage, 2015.
- Skovlund and Peyrot, 2005.** Soren E Skovlund and Mark Peyrot. *The Diabetes Attitudes, Wishes, and Needs (DAWN) program: a new approach to improving outcomes of diabetes care*. Diabetes spectrum, 18(3), 136–142, 2005.
- Spool, May 2004.** Jared M. Spool. *The KJ-Technique. A Group Process for Establishing Priorities*. User interface engineering, 2004. URL https://articles.uie.com/kj_technique/.

- Stanton et al., 2017.** Neville A Stanton, Paul M Salmon, Laura A Rafferty, Guy H Walker, Chris Baber and Daniel P Jenkins. *Human factors methods: a practical guide for engineering and design*. CRC Press, 2017.
- Strauss and Corbin, 1997.** Anselm Strauss and Juliet M Corbin. *Grounded theory in practice*. Sage, 1997.
- Jeanine Turner and James D. Robinson, 2014.* Jeanine Turner and James D. Robinson. Diabetes. In *Encyclopedia of Health Communication*, pages 316–317. SAGE knowledge, 2014.
- Wharton, 1994.** Cathleen Wharton. *The cognitive walkthrough method: A practitioner's guide*. Usability inspection methods, 1994.
- Yin, 2015.** Robert K Yin. *Qualitative research from start to finish*. Guilford Publications, 2015.

List of Figures

2.1	Illustration of how to use multiple samples and plot of the sampling distribution.	7
3.1	Distribution of identified usability problems. The figure shows that no single method were able to identify all usability problems.	18
5.1	Depiction of the full procedure. Each square represents a step in the method. The first step consists of the three analyses. The order of the analysis of the target population and evaluation are indifferent. It is however not possible to map the characteristics' influence in SHERPA without knowledge of both the target population and the HTA.	46
7.1	Depiction of the participants' results from the familiarization. As seen, they are different both content and detail wise. A pink Post It is overall goals, blue Post It's are sub-goals and green are operations. Participant 4 misunderstood the color-significance.	61
7.2	Task analysis for the task "Attach". The pink Post It is the overall goal, blue Post It's are sub-goals and finally the green Post It's are operations. This figure doesn't shows the predicted prerequisites.	62
7.3	Task analysis for the task "Inject". The pink Post It is the overall goal, blue Post It's are sub-goals and finally the green Post It's are operations. This figure doesn't shows the predicted prerequisites.	63
7.4	The added content for the Task analysis of the task "Attach". The sub-goal have been added a operational step.	68
7.5	The additions made for the task analysis of the task 'Inject'. Two sub-goals have been reviewed and nine operations have been added. The sub-goals before are still the same, as the participant developed through the workshop.	69
7.6	The fourth task of the case. Detaching the Device requires to reverse interaction as attaching it.	70
B.1	Depiction of the complete five-phased cycle. The arrows indicate the sequence of the steps. Two-way arrows imply the possibility to move back and forth between two phases, hence making the analysis nonlinear.	104

B.2 Example of different levels of coding. The figure exemplifies how the original textual data can be disassembled and coded into higher conceptual levels (level 1 and level 2). The example is a clipping of a larger table from Yin (2015), where the full example shows the same Category code for several field notes, hence enabling a grouping of categories similar to each other.	105
F.1 Example on why putting all user needs into one design might ruin the experience and functionality for everyone (Cooper et al., 2007).	161
F.2 A depiction of how actual users are mapped to the behavioral variables. This procedure allows for identifying patterns e.g. a cluster of users being mapped similarly across variables.	163
F.3 Typical example of an Identity Model. It contains a description of the identity of the user. Additionally it provides a box with "Give me", which is the needs and goals of this identity.	166
F.4 Depiction of HTA Flowchart for the task "Boil water". It shows how the overall goal can be further split in to sub-goals and operations. It also shows how plans can enable the understanding of sequence.	169
F.5 Depiction of HTA Table for the task "Boil water". It shows how the overall goal can be further split in to sub-goals and operations. It also shows how plans can enable the understanding of sequence.	169
F.6 The typical format for the outcome of a SHERPA. It breaks down all tasks and potential credible errors and provides strategies for remedying or recovering from potential error.	173
G.1 A depiction of the different decisions that can influence the amount of insulin a patient has to take before each meal. Additionally a listing of all the needed physical activities that are required to do, each time. Figure developed by Laagland (2017)	175
G.2 A metaphor for the struggles a diabetic experiences. The metaphor consists of six different barriers that a diabetic experiences. It shows how these six factors makes the patient insecure, unmotivated and frustrated about the disease. Figure developed by Laagland (2017)	177

List of Tables

2.1	Marshall (1996) view on the differences between qualitative and quantitative research.	9
3.1	Distribution of identified usability problems for each type of evaluation	15
3.2	Distribution of identified usability problems visualized. A column represents a usability problem related to the number of participants identifying the particular problem. A black box indicates the identification of a usability problem.	15
3.3	Distribution of identified usability problems for each method	17
5.1	Different types of plans for the Hierarchical Task Analysis (Stanton et al., 2017, p. 50)	41
7.1	Answers for "I liked..." and "I wished..." from the end of the workshop.	56
7.2	The categories who were prioritized by the participants. Only categories which got a score is shown.	58
7.3	A table showing the predicted prerequisites and their affiliated operations, sub-goals and goals of the task. A parentheses () refers to the tag of the post it, which enables back-tracking.	64
7.4	The results obtained by the participants in the development of SHERPA. P (Probability) is questioned, as they did not know. C (Criticality) is denoted by an exclamation mark but is originally judged on an ordinal scale (low, medium, high). The assessment of Sample consequences and Sample Criticality is based on a comparison between the previously predicted prerequisites and the content of this table.	67
7.5	The final result of the task analysis. It includes everything developed through the analysis and states the predicted prerequisites a person would need for completing the individual task, its sub-goals and the associated operations.	71
7.6	All results from the authors use of the methods. The results have been condensed in order to give an overview. From the left-most column is the given sub-goal. Each sub-goals have been assigned a perceptual, cognitive and handling error mode, which is then further broken down. In the last two columns, the consequences for using another sample and the criticality thereof, is described.	74

B.1	A matrix to be used for reassembling data. All entries in the table are invented for the purpose of exemplifying the method, but does suggest how one could see emerging patterns between sample type and method e.g. the first row column showing the same results for first three types of sample. Both numeric and textual data can be used in such matrix, but numeric and textual data doesn't make sense to mix within one matrix, as they are not comparable.	106
C.1	Division of types of sampling by Kothari (2004)	111
C.2	Four authors and their take on sampling methods for qualitative research. While ' <i>outdated</i> ' most of the described methods is still used today (Coyne, 1997).	112
E.1	Meaning condensation from the interview w. Tina Øvad. The <i>Unit</i> is used for reference, hence U1 in the report would refer to the first unit of the analysis. The <i>Theme</i> describes the essence of the natural unit and the <i>Condensed Description</i> is a concise description of the information found in the <i>Natural Unit</i> . The <i>Natural Unit</i> is the actual statement from the interview.	140
E.2	Meaning condensation from the interview w. Morten Purup. The <i>Unit</i> is used for reference, hence U1 in the report would refer to the first unit of the analysis. The <i>Theme</i> describes the essence of the natural unit and the <i>Condensed Description</i> is a concise description of the information found in the <i>Natural Unit</i> . The <i>Natural Unit</i> is the actual statement from the interview.	152
F.1	The error mode taxonomy established in the SHERPA. Each error is referred to by its number e.g., 'operation too long/short' is assigned A1.	171
H.1	Results from the first assignment of the workshop.	217

Appendices

Chapter A

Novo Nordisk and Sampling Difficulties

Novo Nordisk does an astonishing number of evaluations each year, with the goal of increasing the usability of the products and the perceived user experience. They do however also have a big burden of proof, as most, if not all, of their products are sold in USA. This implies that a approval from FDA has to be collected for each product set on the market.

The evaluations are typically conducted in Denmark, UK or USA and most commonly they use samples from the targeted population. Typically they don't have any difficulties sampling from the targeted population, due to their high budgets and resources. Recruiting is however very time-consuming and expensive. Sometimes evaluations are done on employees from Novo Nordisk, who doesn't exactly work with the development of the stimuli being tested. This, of course, saves some of the time and money.

While it is no trouble sampling from the targeted population in cases with diabetes, there seems to be higher difficulties when targeting other more rare diseases.

Chapter B

Assessment of Methods for Qualitative Analysis

When having collected quantitative data, the analysis of it, is often chosen subjectively. There are several methods to use, where some seem identical while others are very different. The two methods that have been chosen to describe in greater detail are somewhat similar regards being structured into steps. Their area of use, however, are not exactly the same, but both still seems applicable for the analysis of the interviews. The content of this appendix is structured by describing, discussing and finally assessing the method. Afterwards the other method is treated the same.

Yin's Five Phased Cycle

Yin (2015) states that the analysis of qualitative data usually is done in five phases. The five phases should help give an overview of the mass of qualitative data collected through a study in some efficient manner. The five steps should be followed like instruction, resulting in a rigorous procedure for the analytic process. Following the steps as an instruction should lead one's ability to draw the conclusions from the study (Yin, 2015, pp. 176-177). While following the five steps it is important to acknowledge that the rigorousness of the analysis is not only obtained by completing the five steps but also derives from precautions. The three precautions highlighted by Yin (2015) are:

1. Checking and rechecking the accuracy of your data;
2. Making your analysis as thorough and complete as possible rather than cutting corners; and
3. Continually acknowledging the unwanted biases imposed by your own values when you are analyzing your data.

Overview

The five steps that one has to go through is depicted in figure B.1. The depiction of the analysis shows the different phases and the interactions between them.

As stated by the depiction, the analysis begins by '*Compiling* a database. Next the data is *Disassembled* and thirdly *Reassembled* again. The two final phases of the analysis are the *Interpretation* of the data and ultimately a *Conclusion* of the study. It is important to remember that the analysis is nonlinear and therefore phases can be recursive. To assess the method provided by Yin (2015) the five phases are described.

Compiling a Database

The phase of compiling data is done in order to organize data in an orderly manner. Organizing the data in a systematic way is most likely to strengthen the outcome of the analysis and leads to more rigorous qualitative research (Yin, 2015, p. 182). The organization of data leads to the compilation of data, which after being processed in the first phase can be considered a database, and not just data in random order. One important aspect of compiling the data is to obtain consistency. Yin (2015) exemplifies this: Through the review of data, one finds that the same terminology has been used differently. This disparate has to be cleared, before proceeding.

The course of action for compiling data isn't universal and the number of different approaches is multiple. Some quantitative approaches may be applicable for compiling the qual-

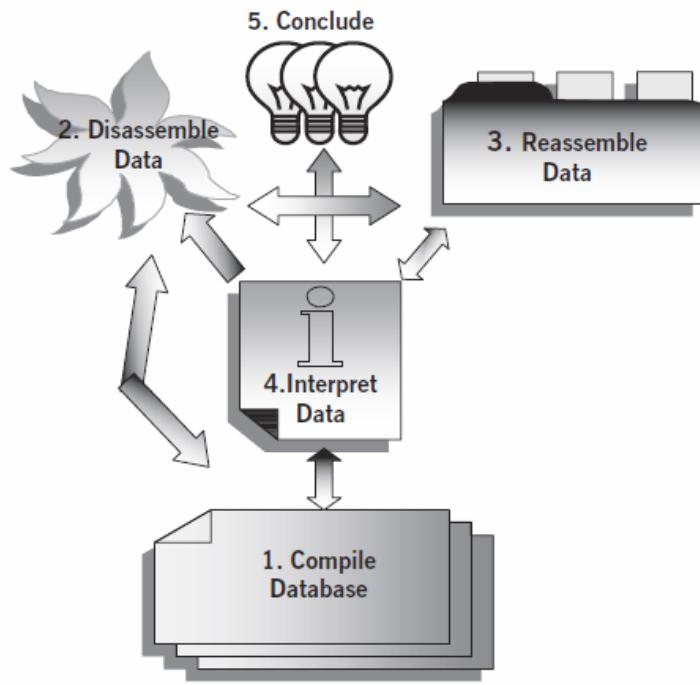


Figure B.1: Depiction of the complete five-phased cycle. The arrows indicate the sequence of the steps. Two-way arrows imply the possibility to move back and forth between two phases, hence making the analysis nonlinear.

itative data, as the main difference between the two is likely to be the type of data. Quantitative data usually consists of numeric entries compared to the textual-based qualitative data, but both approaches strive to order the data in a systematic fashion.

While an orderly, consistent and organized compilation of data is the goal of the first phase, one additional benefit is the familiarization with the data. Through the compiling of data, one has to review the data in order to achieve a compiled database. The review of data happens in less stressed surroundings (compared to being in the field doing the study) allowing for a more thoughtful process of the information collected. Through this process Yin (2015) states that one should consider the following:

- What are the distinctive features of your study?
- How might the collected data relate to the original research questions?
- Are there potentially new insights that have emerged?

Disassembling the Data

Disassembling data is essentially a categorization of data. The categorization can be done on multiple levels and the higher the level of categorization, the higher conceptual level (Yin, 2015, pp. 186-187). An example of this type of categorization is given in figure B.2.

Disassembling data can be approached in several ways as there is no fixed routine for doing

Illustrative words from original field notes	Initial code (Level 1)	Category code (Level 2)
1. "Samantha brought homework home, but she did not always have the right assignment."	Student oversight	Barrier for getting homework done

Figure B.2: Example of different levels of coding. The figure exemplifies how the original textual data can be disassembled and coded into higher conceptual levels (level 1 and level 2). The example is a clipping of a larger table from Yin (2015), where the full example shows the same Category code for several field notes, hence enabling a grouping of categories similar to each other.

so. In general, it has to be decided whether to code or not to code the data. Coding data implies labeling data with respect to the specific content and the contextualized details. The procedure and goal of coding are to methodically move to a higher conceptual level. The coding enables one to sort items (the coded data) into similar or dissimilar groups (Yin, 2015, p. 187).

While the approach of disassembling data by coding can be taken, one also has the opportunity to disassemble data, without the use of coding. Yin (2015) clearly states that the process of disassembling without coding isn't methodological and that the process can lead to non-systematic and inconsistent judgments. To improve the disassembling, without coding, Yin (2015) suggests that the formerly presented precautions are revisited often in order to maintain a rigorous procedure. Even though the disassembling is done without coding researchers has a tendency to disassembling the data without any formal coding, as creative ideas move faster and better with this approach. Yin (2015) finally states that the pitfalls of disassembling without coding can be overcome by revisiting the original data several times and to make sure that the disassembled categories truly represents the original data.

Reassembling the Data

When the data in some way have been categorized patterns and tendencies might emerge, which is the essence of reassembling the data. If one chose to code the data, one could argue that the reassembling of data would be to continue the categorization on higher conceptual levels (i.e. level 3 and level 4) whereby the categories are gathered into themes. If the process of disassembling the data hasn't followed any formal methods of coding, it doesn't have a big impact, if the disassembling was done carefully and the mentioned precautions were taken into account (Yin, 2015, 190-191).

For both scenarios, the important aspect of reassembling the data is questioning the data, as this process is intrinsic for the procedure. While going through the reassembling Yin (2015) mentions four typical question one might ask(Yin, 2015, p. 191):

- Do the patterns make sense?
- Are they moving you to a substantively important plane?
- How do the patterns relate to the concepts and hypotheses of the study?
- Do the patterns become more complicated or expansive when you review your database?

As one continues with the reassembling of data a need for arranging themes differently and altering them might occur. Yin (2015) refers to this as '*playing with the data*'. Additionally, Yin (2015) proposes the use of hierarchical arrays and matrices in order to ease up the task. Arrays are usually used as a sort of scale, where one end-point is tied to more concrete data and the other end is tied to more abstract concepts, derived from the data. This allows for a hierarchical structure that enables one to organize the themes and maybe help to identify associations and relationships between data. Furthermore, matrices can be used for reassembling data, and the most typical way is to define two dimensions of interest and use these for the rows and columns e.g. a comparison of method and type of sample. An example of such a matrix is shown in Table B.1.

	Method 1	Method 2	Method 3	Method 4	Method 5
Sample type 1	A	A	A	B	C
Sample type 2	A	B	C	D	E
Sample type 3	A	B	C	A	A
Sample type 4	B	B	B	E	E
Sample type 5	C	C	E	D	A

Table B.1: A matrix to be used for reassembling data. All entries in the table are invented for the purpose of exemplifying the method, but does suggest how one could see emerging patterns between sample type and method e.g. the first row column showing the same results for first three types of sample. Both numeric and textual data can be used in such matrix, but numeric and textual data doesn't make sense to mix within one matrix, as they are not comparable.

Finally Yin (2015) suggests three procedures to apply while reassembling the data:

1. Make constant comparisons

Watch for similarities or dissimilarities in the data and keep questioning why the data as first was assessed being similar and dissimilar

2. Look for negative instances

Examine everything closely, as some data might have been pooled due to their similarity, but shows not to be. E.g. if one reassembles data into a theme and both positive and negative data are pooled together herein.

3. Apply rival thinking

Search for alternative explanations. One might obtain knowledge at the end of the re-assembling process, that alters the way one would perceive the first items that were reassembled.

With all of the above regarding reassembling in mind, a successful reassembling of data should enable one to identify a broader theme and the outline of the entire analysis. If this isn't the case iterations of the disassembling and reassembling phase can be helpful. If broader themes are present the interpretations and conclusion of data can be done.

Interpreting the Data

Regardless of almost every choice made for a study an interpretation is needed. Therefore, the fourth phase defined by Yin (2015) isn't unique. Instead, some advice is given for reaching a comprehensive interpretation. The main focus on the interpretation shouldn't revolve around data only, but instead, include everything about the study. Encompassing specific data is fine, but the predominantly part of an interpretation should improve the basic understanding of the entirety of the study. To achieve this Yin (2015) lists attributes valuable to consider. He does however still state that no firm definition exists for making a good interpretation. The attributes listed are:

- Completeness
Does the structure of the interpretation include a beginning, a middle and an end?
- Fairness
Would others interpret the results and the study the same?
- Empirical accuracy
Does the interpretation rightfully and respectively represent the data?
- Value-added
Does the interpretation contribute with something new within the field of the study?
- Credibility
Would the interpretation be accepted or critiqued by equals?

While Yin (2015) states the strive for answering the attributes should provide one with a chance of making a comprehensive interpretation, he also presents three 'types' of interpretations. These are listed below, but not described further, as each of them implicitly explains their goal.

- Description
- Description and a call for action
- Explanation

Conclusion on Data

Identical to the interpretation all studies no matter method and analysis should contain a conclusion. Yin (2015) states that the goal of a conclusion is to create an overarching statement which translates the actual findings to broader and higher conceptual levels of ideas.

Furthermore, it should highlight the significance of the study and present the practical implications one has experienced. Based on findings, significance and practical implications one has several choices of conclusion.

If the findings suggest new areas of interest the conclusion could imply the need for additional or new research in areas derived from the study. Another way of concluding is to challenge conventional thinking of certain subjects. Yin (2015) exemplifies this as challenging generalizations and/or social stereotypes. Further, if the findings of the study suggest 'new' information, the conclusion could consist of proposals for new concepts or theories. Finally, one has the options to conclude by making substantive propositions or generalize the findings of the study to other relevant areas.

Assessment

Yin (2015) seems to heavily focus on the identification of patterns in data. While this is a very important aspect it might not be meaningful when analyzing data from single expert interviews, as there might be no pattern. Compared to the expert interviews conducted for the collection of empirical data, this might be true. In addition to the general goal of the five-phased cycle, the coding of data also seems unnecessary, as the interviewee is an expert in the field. The analysis seems very beneficial in situations where multiple interviews have been conducted and are to be compared.

Kvales' Qualitative Analysis

Kvale (2004) divide the analysis of interviews into two categories. The categories are analysis focusing on meaning and analysis focusing on language. The method of analysis presented forward on, all fall into the former category, thus focusing on analyzing the meaning of the interview.

Meaning Condensation

The main principle of meaning condensation is to make sense of longer articulations. This is done by condensing them into short and detailed descriptions holding the essence of what the interviewee mentioned. The true goal of using meaning condensation is to obtain more concise wording of the interview, thus clarifying the main themes of an interview (Kvale, 2004, pp.190-192).

Similar to Yin (2015), the empirical phenomenological analysis used for meaning condensation, requires five steps. The first step is to read through the transcript of the interview(s) in order to obtain an impression of the entirety of the interview. Then the selection of *natural meaning units* is chosen, which is the selection of statements expressed by the interviewee. After this, each natural meaning unit is assigned a theme. The theme should reflect the essence of each unit, hence describing what theme that dominates the particular unit. When all natural meaning units of interest have been assigned a theme, one has to start questioning how this theme relates to the objective of the study i.e. for the interviews it could be "How does this relate to sampling?".

The last and final step of condensing the meaning of an interview is identifying the relevant themes and combine them into a descriptive statement (Kvale, 2004, pp.190-192). A very important aspect of the meaning condensation analysis is that the method isn't limited to phenomenological methods, but can be used in a variety of qualitative studies.

B.0.1 Meaning Categorization

Categorizing the meaning of an interview is by Kvale (2004) said to transpose statements into simple categories e.g. marking with '+' and '-' if a phenomenon is present or not. With this approach, one has the opportunity to reduce one or several interviews into tables and figures containing the key takeaways. The preconceived ease of applying this might be a pitfall, as one has to keep the perspective of the study in mind. This is due to different methods providing different results. Additional focus on the analysis is needed, as one has to know if the analysis leads to qualitative or quantitative data or the objective is to analyze the linguistics or the psychological aspects (Kvale, 2004, p.190).

While Kvale (2004) doesn't describe exactly how to categorize meanings from an interview, he states that coding is essential. The objective of this type of analysis is to identify the main dimensions of a given problem and afterward try to explain the aspects of these dimensions with several subcategories. Kvale (2004) hypothesized that the use of grades influenced both learning and the social relationships in school. Seven main dimensions were decided based on literature and pilot interviews. Through 30 interviews with pupils (762 pages of transcripts), he then identified subcategories for each of the seven dimensions. Each subcategory was defined by a short description, in order to elaborate how a subcategory influenced the dimension.

Assessment

The meaning condensation analysis seems to be applicable to the expert interviews. It seems reasonable due to results to be obtained, as concise descriptions of the different themes might help one to present the data in an orderly manner. This also makes it easier to identify and compare themes across the two expert interviews. Additionally, the main goal of the expert interviews, in this case, is to extract the meaning and look into sampling.

As for the meaning categorization analysis, this is assessed as useful, when dealing with large amounts of explanatory variables for several domains. It seems that in order for this method to be beneficial and useful, one has to obtain a comprehensive set of data on the given subject to enable one to withdraw actual meaning.

Chapter C

Quantitative and Qualitative Approaches

It should be evident that there's some difference between the quantitative and qualitative approaches by now. Before going into detail what these differences mean and how they influence the methods of sampling, the different known methods for sampling are described. The review of existing methods is done to increase the understanding of why such differences exist. Initially, Kothari (2004) divides the two approaches further. The division is shown in Table C.1.

	Probability Sampling	Non-probability Sampling
Unrestricted sampling	Simple Random Sampling	Haphazard sampling or convenience sampling
Restricted Sampling	Complex Random Sampling	Purposive sampling

Table C.1: Division of types of sampling by Kothari (2004)

In the table *unrestricted* and *restricted* is a decision making factor regarding method of sampling. The unrestricted factors describe any situation where a sample from a population with no restrictions apply, i.e., gender, race or special characteristics and the restricted refers to the opposite. While the two terms, to the author's knowledge, doesn't recur in other literature, the described division does however still apply. In the following subsections, it is therefore chosen to present both the qualitative (non-probabilistic sampling) and quantitative (probabilistic sampling).

Methods for Qualitative Sampling

When investigating the existing literature on sampling in qualitative research, a theme emerges. The theme is confusion and misunderstanding, which seems to be caused by the many variations of qualitative methods for sampling. The confusion is due to overlapping methods for sampling, i.e., the names purposeful and theoretical sampling are used synonymously and interchangeable in the literature (Coyne, 1997, p. 623). A clear definition of the difference between sampling methods in quantitative and qualitative research does however exist and is said to be; probability and non-probability sampling, where the former refers to sampling in quantitative research and the latter to qualitative research. The latter is by MacNealy (1999) also described as:

"...[U]sed in circumstances where probability samples cannot be obtained or where levels of confidence are not that important".

(MacNealy, 1999, p. 156)

The quote suggests that using non-probability sampling should be done when the scope of research is not defined to obtain results that are generalizable, but instead to uncover complex issues that can't or at least isn't valuable to be treated statistically, e.g., user experience. Marshall (1996) builds further into this suggestion as he states the importance of recognizing that the meaning of qualitative research is to observe and listen to people in natural settings instead of doing it in a clinical setup. With this, Marshall (1996) reasons that sampling has to take not only the individuals characteristics into consideration, but also account for temporal, spatial and situational influences which together creates the context of the study. Including context forces one to consider biases if the participant's answers are affected by time, place or mental state. If it does not, then great but if it does then alternatives has to be considered regarding the setup of the study. However, Marshall (1996) also mentions that there is no right answer to these questions just as there is no correct way of collecting a sample. Consideration has to be put into these factors as they possibly influence the results, which is valid for all methods of sampling.

Through the revisit of literature, a review of several methods for sampling was conducted. There seems to be many overlaps and variations in these methods.

Strauss & Corbin (1990)	Theoretical Sampling - Three Stages Open sampling Relational and variational sampling Discriminate sampling
Patton (1990)	All sampling is Purposeful Extreme or deviant case sampling Intensity sampling Maximum variation sampling Homogeneous sampling Typical case sampling Stratified purposeful sampling Critical case sampling Snowball sampling Criterion sampling Theory-based or operational construct sampling Confirming and disconfirming cases Opportunistic sampling Purposeful random sampling Sampling politically important samples Convenience sampling
Morse (1991)	Four types Purposeful sampling Nominated sampling Volunteer sampling Total population sampling
Sandelowski et al. (1992)	Selective sampling Theoretical sampling
Sandelowski (1995)	All sampling is purposeful - three kinds Maximum variation Phenomenal variation Theoretical variation

Table C.2: Four authors and their take on sampling methods for qualitative research. While 'outdated' most of the described methods is still used today (Coyne, 1997).

Marshall (1996) states that there might be differences in theory when describing the methods, but in practice, some of the methods will be more alike. The misunderstandings and the confusion with sampling might be clarified by the immense number of methods within non-probabilistic sampling. To visualize the number of different methods a recreation of a table made by Coyne (1997) is shown in Table C.2. It could imply, that sampling is seen very subjectively, as these four authors surely don't have the seem view on sampling.

Due to the variety of sampling methods, it has been chosen to present four method in greater detail. These are:

- Convenience Sampling,
- Purposeful Sampling,
- Theoretical Sampling and finally
- Snowball Sampling.

The choice of these methods is due to their popularity, but also to show why people confuse these methods with each other (purposeful and theoretical). One could argue that four methods are not enough to cover the area of qualitative sampling, but many of the methods given in table C.2 overlap. Some methods overlap, and some can be disregarded to their simple nature, i.e., *Maximum Variation Sampling* which suggests that a variety of users should participate to discover as much as possible. Using a variety of users could apply to all other methods of sampling. Finally, many of the methods seem to have quite specific situations of use.

Convenience Sampling Method

The *Convenience Sampling Method* is what it sounds like, but is sometimes also referred to as the *accidental* or *opportunistic* method in literature, due to the nature of it. This method is the least rigorous, as it revolves around the selection of subject who are accessible. This means that there no guarantee drawing a sample from the targeted population which can result in poor quality data, but it is however very efficient in term of effort, time and money (Marshall, 1996, p. 523) (Koerber and McMichael, 2008, p. 463).

While the above isn't a strong proponent of the use of the method, Koerber and McMichael (2008), defends the use thereof. Some convenience samples might be more accessible than others, but it still takes an effort to reach out and recruit the participants. In addition to this, the convenience and close relationship between researcher, research site and possibly participants can turn out to provide rich data due to the familiarity (Koerber and McMichael, 2008, p. 463).

Koerber and McMichael (2008) does however still point out a critical pitfall, which is that the familiarity can tempt researchers to generalize beyond what is possible, and thereby it is suggested to be careful not to overgeneralize in situations where a convenience sample is used.

Purposeful Sampling Method

The *Purposeful Sampling method* requires, opposite the method of convenience, higher knowledge of the users, both for the targeted population and the substitute population. The requirement of knowledge is due to *purposely* recruiting users for one's sample that matches the necessary characteristics of the targeted population for answering questions about specific subjects most productively (Marshall, 1996, p. 523)(Koerber and McMichael, 2008, p. 464).

Actively selecting the most productive sample can be obtained by developing a sample framework of the variables that influence and contributes the answers. It is suggested to sample a more comprehensive selection of users to have both confirming and dis-confirming users in the sample (Marshall, 1996, p. 523). This suggestion is due to the most obvious pitfall, which is to accidentally sample users who do not represent the diversity of the targeted population. This pitfall could happen if the characteristics defined before sampling are stringent - due to the nature of the method it also becomes easy for researchers to intentionally draw a sample that more likely will achieve wanted results (Koerber and McMichael, 2008, p. 464).

Lastly, the use of Purposeful Sampling has to be appropriately documented. One takes a sample from particular criteria to fulfill a specific scientific purpose; thus a detailed description of procedure has to be in place due to replicating purposes.

Theoretical Sampling Method

Theoretical Sampling can be confused with the *Purposeful Sampling* but in health care literature a clear distinction between the two methods exists (Koerber and McMichael, 2008, p. 465). While the Purposeful Sampling defines the criteria for one's sample in advance, the theoretical sampling method iterates on the initial sampling criteria throughout the study.

In theoretical sampling, some soft criteria are defined to sample from the targeted population, but with the possibility to change, remove or add criteria during the study. It is an iterative process where the criteria for one's sample emerges throughout the study. Additionally, theoretical sampling derives from *Ground Theory*, which has to be considered. The explanation of this is that grounded theory assumes that data collected throughout a study provide and feeds into a new theory. Whenever a phenomenon that can't be explained by existing theory appears, the researcher can change the criteria of sampling to obtain further data on this given phenomenon. Hereafter, the '*new*' theory is changed according to the trends seen in the data. The purpose of using a theoretical sampling method would be to question or confirm the phenomenon, by recruiting samples with characteristics that most likely will be able to do this (Koerber and McMichael, 2008, pp. 465-466).

As one might have figured out, this process of sampling requires amounts of resources significantly higher than the former methods for sampling. Researchers will have to analyze and interpret data each time a new phenomenon appears, to decide the direction of the study and thereby redefining the criteria of the sampling (Koerber and McMichael, 2008, p. 465).

Snowball Sampling Method

Snowball Sampling is the recruitment of a small number of participants, whom one hopes to know other people in the same population they are from that they could recommend. This approach is not limited to specific methods, i.e., snowball sampling can be used in both convenience-, purposeful,- and theoretical sampling as it only requires one to accept having less control of who enters the sample (Marshall, 1996, p. 523).

Snowball sampling is however not useful if the sample needs to be defined and recruited beforehand the conduction of a study, as this counters the goal of using snowball sampling. Even though the method offers less control and requires one to recruit meanwhile conducting the study, it has formerly been used in studies targeting small hard-to-reach population. It proves useful for these - these studies, however, have a list of criteria (drug-users), whereas they seek out a small number of people from this population and ask if they know other similar to them, who could be interested in participating (Platt et al., 2006; Kelly, 2010).

Methods for Quantitative Sampling

When one wants to generalize, and levels of confidence matter, one must pick a quantitative design for the research to be conducted. As with the qualitative research, a variety of sampling methods exists within the quantitative approach. Opposite the qualitative approaches relying on non-probability, the quantitative sampling methods rely on probability sampling, which is also known as '*random sampling*' or '*chance sampling*' (Kothari, 2004).

From such a sample, everything has equal chances of getting picked. It is, however, treated in the analysis of data, where means, probabilities, errors of sampling or significance of results can be calculated and thereby mitigate some of the doubt. Kothari (2004) states that random sampling ensures the law of *Statistical Regularity* which says that if on an average the sample is chosen is a random one, the composition, and characteristics of that sample will be identical to the targeted population. Because of this, random sampling is considered as the best technique when trying to obtain a representative sample.

Due to random sampling being well-documented and arguably the best method for achieving a representative sampling within probabilistic sampling, most, if not all, methods can be further divided into *Simple Random Sampling* or *Complex Random Sampling*.

Simple Random Sample

Simple Random Sample (SRS) can be done when the nature of the targeted population has been defined. Nothing else has to be prepared or defined, as users from the targeted population simply can be selected and asked for participation. One important aspect of SRS is that all members from the targeted population have an equal chance of selection, i.e., no control on varying aspects in the targeted population as gender, race or age (Marshall, 1996, p. 1).

Stratified Sampling

Variants of the SRS exists, which extends to and includes other aspects; thus it is possible to study subgroups from the targeted population in greater detail, e.g., gender, race, and age which the SRS does not take into account. These additional criteria are what increases the complexity of the sampling.

Using the technique of stratified sampling requires dividing the targeted population into smaller subgroups. A criterion for making a subgroup is that the subgroup should be more homogeneous than the overall targeted population, e.g., if one suspects gender to have high influence, a subgroup of men and a subgroup of women would be more homogeneous than the two combined. When the division of the targeted population is considered done, the SRS is then used on each of the subgroups created and finally combined into one sample, which

is a stratified sample. One other addition is that the number of participants drawn from each subgroup should be proportional to their actual frequency of the targeted population before it is self-representing. If this is not taken into account another way of obtaining this, would be to weight each subgroup (Rivers, 2008, p. 6).

Systematic Sampling

The sound and wording of *Systematic Sampling* doesn't exactly lead one's thoughts onto random sampling but is treated alike in literature. The typical procedure of systematic sampling is to have a list available of potential subjects. When achieving this, the desired sampling-percentage is determined.

Choosing the percentage can be done by desire, e.g., if 10 percent has been determined, one randomly picks one of the first ten entries on the list. Next, systematically take every 10th after that. E.g., if choosing the fourth item on the list, the next pick would be the 14th element in the list.

The systematic approach has the advantage of sampling more evenly spread out one's targeted population and is described as convenient and cost-effective compared to the regular random sampling when a sizeable targeted population is available (Kothari, 2004, p. 62).

As with other methods, systematic sampling cannot be used under certain conditions. One of these is if the listed population has a hidden periodicity one will experience a high sampling error hence being inefficient as it will not be able to represent the population truly. One possible solution to this would be to randomize the entries in the list as this would make systematic sampling to be considered equivalent to random sampling in general.

Chapter D

Transcripts of Interview

The following appendix contains the transcription of the interview conducted with Tina Øvad and Morten Purup. The goal of transcribing is to obtain a text, which can be read and understood. Therefore, all indications related to aural information are omitted.

Transcript of Interview with Tina Øvad

- 1 **F:** ... Heldigt at vi kan snakke sammen, og jeg er glad for at du gider tage dig tid til det.
- 2 **P:** Ja, altså det lyder vildt spændende...
- 3 **F:** Ja, altså jeg ved ikke, vil du have lidt baggrund på det, eller...
- 4 **P:** Meget gerne!
- 5 **F:** Altså, man kan sige jeg har brugt et halvt år ude i Novo Nordisk, i praktik, hvor jeg ligesom har lavet usability
- 6 og user research. Jeg så ligesom en mulighed i enten at spare nogle ressourcer, i forbindelse med det her rekrut-
- 7 tering og sampling af brugere, men jeg så også den udfordring, at når det er medicin som måske er til sjældnerte
- 8 sygdomme, som også udvikles derude, så kan det være utrolig svært at få fat i den der population.
- 9 **P:** Ja....
- 10 **F:** så hele mit speciale, går egentlig lidt ud på hvordan man ligesom tilgår sampling, når ens population ikke er i
- 11 rækkevidde eller hvis ens produkt er så konfidentielt, at det ikke kan komme ud af virksomhedsrammerne. Min
- 12 vejleder nævnte så at du arbejder med noget, der var lidt lige sådan.
- 13 **P:** Ja, altså det er.. Det er ikke helt så hårde omstændigheder, som du arbejder i, må jeg erkende. Men lidt på
- 14 siden, er jeg i hvert fald i gang med at kigge på det her med, hvornår tester vi med hvilke typer folk, ikke? Fordi vi
- 15 har jo, i forhold til vores arbejde at øhh - altså nogle gange bliver vi nød til at teste med potentielle end-customers,
- 16 men nogen gange er det jo også human factors vi tester, nogen gange er det sådan noget lidt mere overordnet,
- 17 hvor vi ligesom kan sige jamen altså det er ikke super relevant det lige er en ende-bruger der er med her og så er
- 18 der især, hvis vi taler selve UX-delen af det, i forhold til os, ikke? Der bliver vi som regelt nød til at skulle have
- 19 nogle fra vores kunde-segment ind og teste med det.
- 20 **F:** Ja, det er lidt der... Jeg har jo så læst helt basiske teorier om sampling også ligesom også prøvet at finde
- 21 noget empiri gennem studier. Jeg synes ligesom der er den her opdeling af at enten bruger man sine end-users
- 22 eller også bruger man nogle eksperter - og der er ikke rigtigt noget litteratur på, hvad man sådan kan gøre, hvis
- 23 populationen er uopnåelig.
- 24 **P:** Ja... ja fordi det kommer også lidt an på ikke, fordi jeg sidder lige og tænker i forhold til du siger det her med
- 25 at det er medicin til sjældne sygdomme, ikke?
- 26 **F:** Ja, det kunne være en case, ja.
- 27 **P:** Ja, for eksempel ikke - eller det er netop det her med en population uden for rækkevidde, men jeg tænker jo
- 28 stadigvæk, altså hvor meget er det der er noget der er hængt op på den specifikke sygdom, altså hvor det er et eller
- 29 andet der gør at de skal gøre tingene på en anden måde end man normalt ville gøre.
- 30 **F:** Ja, ligenøjagtigt. Det er den... Den løsning jeg har en hypotese om der kunne fungere, det vil være det her
- 31 med at man går ind og analyserer hvad der egentlig er brug for i den enkelte test
- 32 **P:** Lige præcis.
- 33 For hvis det er håndteringsmæssigt så er det jo meget... altså medmindre der er nogle perceptuelle udfordringer
- 34 ved den sygdom, så burde almindelige mennesker i realiteten kunne [håndtere...
- 35 **P:** [Lige præcis
- 36 **F:** Så det er lidt hypotesen, men jeg har ikke rigtigt noget som helst der støtter op om det, andet end sund fornuft-

37 agtigt.

38 **P:** Ja, jeg må jo så også erkende at jeg slet ikke har kigget ind i det, fra et teoretisk perspektiv. Altså jeg har
39 kigget ind i det fra et meget meget praktisk synpunkt der hedder - jamen altså vi kan jo gå ind og vurdere; er
40 det human factors, så kan vi teste med mere eller mindre hvem som helst; er det UX-delen af det så tester vi med
41 potentielle kunder, ikke? også er der et eller andet range in between, hvor man så bliver nød til at graduere lidt,
42 ikke? Fordi jeg tænker sådan lidt, altså umiddelbart i forhold til dit - altså jeg jeg bliver ved med, når du taler
43 om det, jeg tænker sådan en eller anden matrix op, ikke? Hvor det er sådan noget med at i forhold til human
44 factors, så er det den her type ting vi kan teste og det er egentlig lidt ligemeget hvem vi har inde, det er bare
45 nogen hvor der ikke er nogen dissabilities også videre også videre. Hvis at det er at det er noget, hvor der er
46 nogle dissabilities, vil man så stadigvæk kunne finde sit sample i en anden gruppe end dem der lige præcis har
47 den specifikke sygdom? I min optik, der tror jeg... Altså jeg hører det meget som at det kan skiller rigtig meget
48 ad. Altså det her med, at det kan godt være det er en sjælden sygdom, men hvad impact har det på patienten?
49 Hvis det ikke er noget der har impact på, netop som du også selv siger, med noget kognitivt eller noget fysisk
50 af en eller anden art, ikke? Så umiddelbart tænker jeg det kunne være interessant at kigge på - hvordan skal jeg
51 formulere det - en form, altså det er jo ikke en persona, men det er jo lidt derhenaf, ikke? Ligesom så at sige, du
52 har den her sygdom, det kan være X, [ikke?

53 **F:** [Ja

54 **P:** og i forhold til X, der er der nogle kriterier det kan være at personen er lam i hænderne for eksempel, ikke?
55 Eller det kan være... Jeg ved ikke hvad det kunne være, men det kan være mange forskellige ting også simpelthen
56 stille de der kriterier op, der vil være ved de forskellige sygdomme, som kan have indvirkning på en interaktion.
57 Også simpelthen caste... rekruttere efter det, ikke?

58 **F:** Jo, fordi... Lidt tilbage til det der med håndteringen, det er jo lidt at hvis sygdommen - lad os sige forårsager
59 ledigt i fingrene, så kan man jo måske godt substituere med folk. Altså bare kigge på folk der har ledigt i
60 fingrene, hvis man udelukkende fokuserer på håndtering.

61 **P:** Lige præcis. Så det er vel lidt det der med de forskellige faktorer der egentlig spiller ind, ikke? Altså, fordi så
62 vil det være - ligesom at kunne sige i forhold til Sygdom A her, så kan det være ledigt, ikke?

63 **F:** Ja

64 **P:** Det kan så have indvirkning på håndteringen af, eksempelvis en kanyle, hvis vi nu siger det er noget allá det
65 i laver, ikke? Men er der andet der så også vil have indvirkning i forhold til den sygdom A? Altså er der noget
66 andet i den profil af den sygdom, der gør et eller andet, ikke? Så kan det være, jamen at det også medbringer
67 farveblindhed, så man skal sørge for at man tester for farveblindhed også. Simpelthen tage kriterierne hele vejen
68 ned og få lavet sådan et kort over de forskellige. Så kan vi sige, i forhold til ledigt, så kan vi jo højest sandsyn-
69 ligt rekruttere alle folk med ledigt, vi behøver ikke folk med en specifik sygdom, vi skal bare have nogen der
70 har ledigt. Vi kan også gå en gang længere ud - jamen hvad er ledigt så? Er det et eller andet med at det er
71 hænderne der ikke kan bevæge sig, men er der andre personer eller andre sygdomme, der medfører det samme,
72 som vi faktisk så reelt set vil kunne teste på den samme type i den kategori.

73 **F:** Ja og det føder fuldstænding ind i det jeg har indtil videre og man kan så sige det næste problem det er så så
74 snart du går over, hvor det faktisk er de her lidt mere kognitive faktorer der gør sig gældende - jeg har rigtigt
75 svært ved at se en mulighed i hvordan man kan substituere den her population, hvis man ligesom er ude efter de

- 76 her lidt mere User Experience relaterede ting, hvor det ikke bare er håndtering og usability.
- 77 **P:** Ja, men spørgsmålet er om man så skulle tage og vende den på hovedet også sige "okay" der er nogle kognitive
78 ting der spiller ind, så de her personer vil ikke kunne x, y og z også vende den den anden vej rundt også sige vi
79 ved de ikke kan x, y og z, hvordan er det så vi sørger for at de ikke har behov for at gøre x, y og z.
- 80 **F:** Ja, så måske [...] hvad skal man sige
- 81 **P:** [Så i stedet for egentlig og hjælpe med at gøre x,y og z så skal man egentlig gøre sådan at hjernen ikke
82 behøver at gøre x,y og z eller i hvert fald minimere x,y og z.
- 83 **F:** Ja, jamen helt sikkert, men jeg tænker også den retning der ligesom hedder i forhold til at når du bedømmer
84 user experience så, i hvert fald inden for diabetikerenes verden, der har de jo typisk været knyttet til et produkt
85 dagligt, henover en årrække. Når de så får noget nyt, der har de jo rigtig meget med i bagagen, i forhold til
86 hvordan de bedømmer det nye.
- 87 **P:** Ja
- 88 **F:** Og sådan som jeg ser det så er den eneste måde og gå ind og substituere noget som helst der, det vil være at
89 man tager sig nogle forbehold for, hvilke begrænsninger der er, ved ikke at bruge de rigtige brugere.
- 90 **P:** Ja... dut dut dut, jeg sidder sådan lidt og tænker, fordi det er netop det der med at hvis der er rigtigt meget
91 læring inde, netop som ved diabetikere, og i forhold til de her insulin penne og de er vant til at stikke sig selv
92 og det er lige ud af landevejen - men man kan også vende den om og sige, hvis man nu laver noget til den her
93 målgruppe, hvor det simpelthen bare er ekstremt nemt og let at have med at gøre?
- 94 **F:** Ja?
- 95 **P:** Fordi man kan jo sige... Man kan jo gå langt ud, og man kan gå den der vej, hvor man gør det så simpelt at
96 hvis du har noget erfaring med det er det piece-of-cake, men er egentlig reelt set også piece-of-cake, selvom du
97 ikke har det. Det er jo så sværere for jer som virksomhed at gøre det, ikke?
- 98 **F:** Jo, og [...]
- 99 **P:** [Vi ved jo at de har en eller anden form læring allerde, så der er allerede noget vi kan tage for givet, at det
100 behøver vi ikke at have klaret altså med udmåling af hvor meget de skal have eller hvor ofte de skal tage det eller
101 hvad det nu kunne være, ikke?]
- 102 **F:** Ja, så lidt de der meget subjektive faktorer... Hvis man lige som kan bevise at det er nemt at gøre det.
- 103 **P:** Ja, fordi det tror jeg i hvert fald... Det er den vej jeg ville gå, altså hvis der virkelig ikke rigtigt er andet, så
104 kan vi gøre det så simpelt at ligemeget hvem det er, ville kunne gøre det også vil det bare være en fordel, hvis de
105 faktisk har erfaringer fra lignende produkter.
- 106 **F:** Ja, men hvad tænker du så, hvis det er et nyt koncept, og der ikke rigtigt er noget data på om det er simpelt
107 eller ej? Fordi det kræver meget af designet at man bare kan lave et statement [om...]
- 108 **P:** [Helt klart, og det bliver nogle iterationer man skal ud i og virkelig prøve det af. Og det kommer virkelig
109 an på det koncept man arbejder med, tænker jeg og hvor det er man er på vej hen. Hvis det er et produkt der...
110 Man bliver nød til at starte et sted og nød til at ligge en baseline der siger hvor simpelt det skal så også være og
111 hvor meget kan vi egentlig regne med der er af læring allerede i den setting. Som jeg ser det, så bliver det sådan
112 noget mapping, hvor man virkelig får det mappet ud. Altså så siger man, vi har produkt A her, og vi ved at vores
113 kerne-kunder de har den her erfaring også har de de her problemstillinger i forhold til, eksempelvis ledigt som
114 du nævner, ikke? Altså simpelthen bare få mappet ud alt det der egentlig kunne have indvirkning også egentlig

115 få lavet en fin sti, der så dækker så meget som muligt. Jeg tænker virkelig det er sådan noget med at man har en
116 eller anden form for matrix, hvor man kan skravere noget af, også sige altså det her er dækket ved at teste sådan
117 her, det her er dækket af at teste den her type personer, det her er dækket ved at teste dadadada også egentlig
118 sige alting er testet, også summere op at det hele er testet, men så måske ikke med den specifikke, hvor det er
119 den samme person der har kunne dække alle de krav, der nu en gang er.

120 **F:** Ja, så lidt det... sådan som jeg forstår det, med sådan en matrice her, at du måske får lidt flere iterationer,
121 fordi du er nød til at have nogle forskellige.

122 **P:** Ja, det tænker jeg. Så simpelthen gå ind og sige nu er det alt det her med eksempelvis håndtering, så det
123 kan være at man bliver nød til at teste med ledig - så kan du sige i forhold til den fysiske interaktion, så er det
124 egentlig dækket i forhold til den problemstilling. Så kan du sige der er nogle kognitive faktorer der spiller ind,
125 jamen så enten lave det så nemt at det virkelig bare er lige ud af landevejen eller finde nogle der har nogle af de
126 samme problemerstillinger, også kan du sige, så er det testet ved case 2. Så har du noget andet, hvor du så kan
127 teste det ved case 3, også har du noget fjerde, hvor du så kan teste det på en eller anden måde også. Så nu er
128 det alt sammen testet, men det er ikke med den samme person og samme målgruppe, men så må vi egentlig bare
129 konkludere at fordi det faktisk virker i forhold til folk der har den her problemstilling, den her problemstilling
130 og den her problemstilling så må vi kunne konkludere at produktet faktisk også vil kunne virke til dem det er
131 designet til.

132 **F:** Ja, okay.

133 **P:** Så jeg tror virkelig det sådan umiddelbart, vil jeg tænke, det vil være sådan et kort man tegner over det - få
134 testet så meget som muligt, men så bare gøre det i forskellige steps, ikke?

135 **F:** Jo, så studiet... Hvis man havde et studie til de rigtige brugere og sammenlignede med det man ville lave, så
136 ville det man ville lave, det ville være mindre studier

137 **P:** Ja, også måske 5 af dem, men med forskellige 'foci' i hver.

138 **F:** I forhold til det her med, hvis de har erfaringer og læringer - sådan noget som kognitive walkthroughs og
139 sådan noget, hvor man ligesom giver noget konteksts og selv sætter op, hvordan det burde være - tænker du at
140 yderligere information til en bruger, som måske ikke kommer fra den rigtige population, at det ligesom vil være
141 med til at give noget?

142 **P:** Ehhhhhmmmm... Jamen jeg tænker næsten det modsatte, altså hvis jeg forstår de spørgsmål rigtigt? Fordi
143 jeg tænker netop at der har i behov for at pumpe dem for så meget information som muligt?

144 **F:** Ja, jeg tænker mere, eksempelvis, hvis jeg hyrer en helt naiv, som aldrig har haft med diabetes at gøres. Hvis
145 jeg så fortæller ham en diabetiker vil typisk gøre det her 3-5 gange om dagen, tror du det vil give noget input,
146 som ligesom er bundet lidt sammen med en erfaring, som en rigtig diabetiker vil have?

147 **P:** Den synes jeg er svær, for et eller andet sted er jeg delt. Et eller andet sted tænker jeg det vil være en fin
148 kontekst at sætte omkring tingene. Jeg ville være lidt bekymret for at jeg ville komme til at farve dem for meget,
149 fordi jeg tror der vil være forskel på - for eksempel med diabetiker-eksemplet, der er du så vant til at du stikker
150 dig fx. 5 gange om dagen altså hvor du ikke tænker over det, men hvis du siger det til mig, så ville jeg tænke
151 "Hold da op, det er godt nok mange gange" og hold da op, så ville jeg have mange forbehold. Så jeg tror det er
152 meget vigtigt det der med, at gå ind i hver også lige tage det der reality-check, hvordan ville det påvirke en at
153 høre det, i forhold til en der er vant til det. Jeg tænker så også til gengæld, at nogle gange tror jeg måske godt

- 154 det kunne give mening og gøre det. Men jeg tror virkelig det er en vurderingssag fra case til case, fordi jeg kan
155 se nogle fordele, men jeg kan altså også se nogle ulemper.
- 156 F: Ja, jamen helt sikkert. Jeg ser også ulempene, men det er lige så meget det her med at fokus er at substituerer
157 den her population, som man ikke kan nå, ligemeget hvad. Og det er sådan lidt, i mit hoved, det er det her med
158 at man måske godt gøre nogle ting, eksempelvis give noget konteksts [...]
- 159 P: Ja, så sætte scenen for det?
- 160 F: Jo, men så samtidig så skal man også have i baghovedet, hvilke begrænsninger det ligesom giver.
- 161 P: Ja. Og ja, jeg tror også du kan komme rigtig langt, ved netop når du har resultaterne, så sætte rammen
162 omkring det og sige det er i den her konteksts det er blevet talt om og det er den her type information de har
163 fået. Så begynder man også at kunne tegne sig lidt et billede af hvor man er på vej hen med det. Så jeg tror
164 det er meget vigtigt det der med at kunne sætte den rigtige konteksts, eksempelvis hvis vi nu siger det her med
165 stik-eksemplet igen, det er med folk der ikke er vant til at stikke sig og de har fået den her information og det
166 er det her resultat der er kommet ud af det. Der er meget i analyse-arbejdet man skal tage højde for når man så
167 analyserer det, ikke?
- 168 F: Ja, så man kan måske argumentere for, at selvom det ikke kommer fra den rigtige population, så hvis der
169 bliver taget hånd om det, både før og efter [...]
- 170 P: [Ja og stadigvæk have i mente, det her med at man også vil kunne farve dem i en retning af de ting man siger,
171 hvis det er noget de ikke er vant til, hvor den population du egentlig gerne ville have testet på det er noget der
172 egentlig er en del af deres dagligdag og deres daglige vaner.
- 173 F: Ja, det er selvfølgelig rigtigt, det er bare også utroligt komplekst
- 174 P: Ja, det er ekstremt kompleksts! Jeg tænker også stik-eksemplet er et virkelig godt eksempel, for jeg tror en
175 diabetiker vil tænke "Nåh ja", de er vant til det og det er en del af deres rutine, så det vil jeg slet ikke tænke over.
176 Hvorimod, hvis du ikke er vant til det eller endda har nålefobi, eller hvad det nu er, så vil man nok tænke "shit
177 mand - stik sig selv 5 gange".
- 178 F: Ja, og det er jo lidt de udfordringer, man sidder med tilbage. Når man sådan kigger tilbage på teorien og
179 de studier der er lavet, så synes jeg bare det hele siger lidt, det er vigtigt at have den rigtige population, men
180 samtidig siger det også lidt, hvis du begynder at sammenligne studie-design, eksempelvis field-evaluations og
181 laboratorie-studier, at man kan godt afdække noget af det samme, men man skal igen bare være klar over hvad
182 betydning kontekstten har, for det man laver.
- 183 P: Lige præcis, ja - og det er det virkelig. Det er også derfor jeg tænker at det mest brugbare vil netop være det her
184 med... Virkelig lave det her... Altså et eller andet skema over... Okay den her target-group, hvad kendetegner
185 den? Også ligesom begynde at dele ind; de har den her problemstilling - hvordan kan vi teste det og hvad er en
186 alternativ rekruttering til det? I forhold til det her med alternativ - det er altså virkelig at få det delt op så man
187 maksimere, hvor mange man potentielt vil kunne bruge i forhold til det specifikke studie.
- 188 F: Ja, altså sådan nogle identitetsroller?
- 189 P: Ja, det er ligefør jeg tænker personaer, fordi i stedet for at have en person, så er det egentlig en sygdom. Så
190 kan man sige, hvis det er diabetes, hvad er diabetes; de skal i hvert fald have noget insulin. Så er det nogen gange
191 skal vi teste med nogen der kan stikke - altså nogen der er vant til at stikke sig selv, ikke? Også hvad det nu ellers
192 kunne være med diabetes, ikke? Simpelthen sige hvad kendetegner det og hvor vil vi kunne få den samme type

193 person? Nej, ikke samme type person, men den samme øhhh.... Hhvad kalder man det?

194 F: De samme karakteristika der beskriver... [

195 P: Lige præcis, ja lige præcis.

196 F: Ja, og så er det jo ligesom at arbejde sig frem til, hvad de her karakteristika er, og det er jo så både de per-
197 ceptuelle og de kognitive.

198 P: Ja lige præcis, og så finde alternativer til hver - også vil du højest sandsynligt støde på at der er noget du
199 ikke vil kunne teste. Altså hvor du så netop så skal ind og sige "Nu mener vi at vi har gjort det så simpelt, at
200 det burde kunne lade sig gøre", men hvor man reelt set ikke vil kunne validere det, medmindre du så får den
201 specifikke gruppe ind. Der vil jeg så tænke at alle iterationerne, så prøv og se om du kan gøre det med nogen
202 der... Det kommer lidt an på, for det kan også være man skal starte op med faktisk at have nogen inde, der har den
203 specifikke sygdom, så man virkelig får noget føeling med, hvor de er henne, ikke? Men i hvert fald være meget
204 påpasselig med hvornår mener du at det er validt at kunne teste med de andre og hvornår mener du så at det
205 virkelig er vigtigt at du tester med den specifikke målgruppe.

206 F: Man kan sige, at nu har jeg selvfolgelig gjort det lidt svært for mig selv, ved bare at sige at den egentlig pop-
207 ulation er utilgængelig. Man kan selvfolgelig sige, at hvis man bare kan udskyde, sådan at man måske skal teste
208 på dem én gang i stedet for 10 gange, det er jo også...

209 P: Det er jo også... Ja det er bedre - meget bedre end ingenting og meget bedre end at skulle gøre det hver gang.

210 F: Ja, lige nøjagtig.

211 P: Ja, for jeg tror virkelig det der med at få mappet det ud og ligesom kunne sige - der er nogle steder, hvor der er
212 et alternativ - og når du så får det mappet ud, så er der også hvor du simpelthen ikke kan finde noget alternativ.
213 Så må du så sige at enten så kan vi komme over det ved og så gøre et eller andet, hvor vi, med de kompetencer
214 vi har, mener at det er klaret eller at der bliver du simpelthen nød til at - lige der - bliver du simpelthen nød til at
215 skulle teste for at kunne være sikker eller også må det være 'Wild Carded' der er lige i den del af det, ikke?

216 F: Jo, og hvad skal man sige... Det der Wild Card eller de alternativer hvor man ikke kan teste, det vil jo så igen
217 være situationsbestemt af eksempelvis diabetikere, de har måske nogle ting som er uopnåelige, men så snart du
218 ligesom kigger på en anden population til et andet produkt, så er det måske noget andet?

219 P: Ja

220 F: Så det her med at lave et generelt værktøj i forhold til at udvælge samples, det synes også lidt uhåndgribeligt,
221 for det kan jo skifte fra gang til gang.

222 P: Jamen præcis.

223 F: Man kan sige [NAVN] har en idé om, at slutresultatet skal være sådan et træ, hvor man kan starte med at
224 spørge "Er det kvantitativt, er det kvalitativt?" også længere og længere ned, indtil du rammer den metode man
225 egentlig vil have også skal der være et forslag til en population. Det skal selvfolgelig ikke være forslag til den
226 specifikke ting man skal substituerer, men måske nogle råd til hvordan man finder alternative mennesker til den
227 her sample.

228 P: Et eller andet sted giver det meget god mening - men mener du at de folk, der skal bruge det her, vil de have
229 kompetencerne til at sige lige nu står jeg i en situation, hvor at det er... hvor jeg har mulighed for... Det er lidt
230 det det igen med vi står med en mulighed for at kunne lave noget kvalitativt eller kvantitativt, hvor jeg tænker
231 jeres, det lyder bare lidt som om det er vendt lidt mere på hovedet, også siger man "Vi har en problemstilling her

- 232 vi skal have testet også har vi mulighed for at gøre lige præcis hvad der er behov for i forhold til den kontekts.
- 233 **F:** Øh, lige i forhold til diabetikere, så er det ikke så stort et problem, fordi dem er der ret mange af, men så snart
- 234 man ligesom begynder at gå udover diabetes og bevæger sig over i mere sjældne sygdomme, så er de spredt for
- 235 alle vinde, der kan du ikke bare rejse til en stat og ligesom få fyldt op.
- 236 **P:** Nej, lige præcis og det er også derfor jeg tænker lidt, at det der med et træ, det giver mening nogen gange.
- 237 Hvor man siger, lige nu der har vi mulighed for at få tre personer ind, nå, men når vi har tre personer, så bliver
- 238 det nok noget kvalitativt det er nok der vi får mest ud af det for vi kan ikke rigtig sige noget kvantitativt om
- 239 noget. Nå, men okay så har vi mulighed for at have dem i en halv time, fordi der er nogle kognitive faktorer der
- 240 spiller ind over et eller andet, lalalala den vej rundt. Jeg tænker lidt jeres er vendt lidt på hovedet, ikke, hvor jeg
- 241 tænker, umiddelbart som jeg hører det, at i står mere med... I har et produkt og står med en problemstilling også
- 242 er det den vej man tager den ind.
- 243 **F:** Øh, ja prøv lige og uddyb...
- 244 **P:** Hvor I så vil sige, I har det her produkt, og nu har I behov for at få testet et eller andet omkring det her
- 245 produkt, hvor I så kan sige, vi har produktet også ved vi at vores target group ser sådan her ud, hvor vi måske
- 246 vil have mulighed for at teste en gang med dem, men der er så alt det her andet vi også gerne vil have testet i
- 247 forhold til kognition, det her med at stikke, i forhold til whatever. Jeg hører det lidt den vej ind, altså hvor i har
- 248 mere mulighed for at sætte.. jeg ved ikke om jeg kan formulere det rigtigt.. Det er mere det her med om I har
- 249 personerne også kan sætte testen op, eller om I har testen også finder personerne bagefter.
- 250 **F:** Ja, og man kan sige, set fra Novo-perspektiv så synes ressourcer i forhold til bare at have testen også finde
- 251 personerne til at være rimelig ladesiggørligt, men det erude i de der cases, hvor det ikke kan lade sig gøre, at det
- 252 begynder at være problematisk og det er ligesom der hvor man gerne vil have en substitute.
- 253 **P:** Ja, præcis fordi det er netop der.. Jeg bliver nemlig lidt triggered af det der udvælgelsestræ, ikke?
- 254 **F:** Ja, det er jo et forslag til hvad specialet eller hvad hele studiet egentlig skal ende ud i, og jeg har da også
- 255 haft mine bekymringer: Ét, det bliver meget specifikt, hvis man vælger et fagområde så er det måske svært at
- 256 generaliserer til noget som helst og to, at verificere at det man ligesom stiller op at det giver mening. Der skal du
- 257 jo have en eller anden form for verificerende test for hvert udsagn du giver.
- 258 **P:** Lige præcis, og jeg tror altså også at jeg er stødt på sådan et tool der en gang. Noget webbaseret, hvor man
- 259 netop siger det er de her omstændigheder jeg arbejder indenfor, hvilken metode skal jeg så vælge. Jeg synes jo
- 260 netop det er selve rekrutteringen og det her med at afdække hvordan man tester de forskellige requirements, det
- 261 synes jeg jo lyder sindsygt spændende. Det andet der, det er sådan lidt en gammel traver, ikke? Men det der
- 262 med hvordan mapper man ud, i forhold til at skulle teste et helt koncept, hvor man bliver nød til at dele det op i
- 263 mindre tests for til sidst at kunne sige vi har valideret det store koncept med 5 mindre koncepter - så mener vi
- 264 bestemt også at vi kan validere hele projektet, ikke?
- 265 **F:** Ja, jamen det giver god mening.
- 266 **P:** Ja det er virkelig spændende synes jeg.
- 267 **F:** Nu ved jeg ikke, men hvor meget tid har du? For jeg kunne også godt tænke mig at vende det her med at hvis
- 268 man har et produkt der er færdigt - eller har sådan et nogenlunde færdigt udseende-mæssigt, så lyder det som
- 269 om at man lægger et sort klæde over også er adgangen til det ikke ret høj.
- 270 **P:** Nej, det er vi ved at gøre op med - rigtig meget op med. Altså jeg mener der skal meget øje på... Det kommer

271 jo an på, nu spørger du jo mig, ikke? Jeg siger jeg mener der skal så mange øjne på som muligt, også bliver vi nød
272 til at sortere i de udsagn der kommer, og ligesom have den professionalisme der siger, altså i forhold til dem der
273 udtales noget, hvilken baggrund har de og hvorfor bliver der sagt som der gør. Der er så også dem der siger vi
274 skal ikke vise det til for mange mennesker, for de er for farvede og lalalala. Jeg er helt klart af den overbevisning
275 at jo flere inputs du kan få, jo bedre, men du skal så også være god til at sortere i de inputs du får. Og altså jeg er
276 helt klart fortaler for, at hvis du har noget, har billeder på din telefon eller et eller andet, hvis du har mulighed
277 for at vise det til nogen, måske in-house, der har en NDA osv. ikke, for at få noget input, men selvfølgelig med
278 det forbehold, der hedder at man skal tænke på, hvem de er og hvor de sidder hen og hvilken baggrund der er.
279 Men der er altid - jeg synes der kommer mere positivt ud af det end der kommer negativt ud af det, i sidste ende.
280 Om ikke andet, det gør dig skarp på at forklare konceptet og det gør også dig skarp på så ligesom at sige... Jamen
281 nogen gange bliver der sagt noget der kan trigge dig. Det kan godt være det ikke er det specifikke der bliver sagt,
282 men det er mere en eller anden lille kommentar der kommer, hvor du kan tage det op og sige "Gud, det har jeg
283 slet ikke tænkt på, men gud der er da et eller andet her jeg bliver nød til at grave lidt mere ned i og lige få kigget
284 på". Så altså jeg er meget stor fortaler for, ja... At vise det til så mange som muligt.

285 **F:** Der kan man jo så sige, at det er lidt at arbejde sig væk fra det der med at det er så konfidentielt at man ikke
286 kan vise det uden for virksomhedsrammerne

287 **P:** Ja, og det er virkelig noget vi er ved at gøre op med. Vi havde vores første sådan brugertest i London sidste
288 uge, hvor det er første gang vi selv er ude og lave det, altså med interne folk og hvor vi gør det selv. Vi har
289 været ude og lave nogle småting med få personer, men her var det med 20 personer, - nej 18 personer til to
290 koncepter, i forhold til portofolio og så videre. Bare det her med at få vist de prototyper der nu en gang er. Jeg
291 er så til gengæld ikke fortaler for det her med high-fidelity prototyper, det har vi sådan en tendens til stadigvæk
292 at gøre rigtig meget i. Men jeg kan godt lide det her med, at man har sådan nogle idéer, nogle tegninger eller
293 nogle renderinger eller et eller andet, så man så kan snakke ud fra det og ligesom bare høre hvad folk har at sige.
294 Jeg har også den der meget med at jeg går meget ind for informeret design og informerede beslutninger, men
295 det kan godt være vi ikke beslutter det folk siger, altså også når vi laver brugertests, det kan godt være at folk
296 de siger de vil have - altså vi viser dem A, men så siger de, de vil have B, men så kan vi godt gå ind nogle gange
297 og sige, det kan godt være de siger det, men vi vælger faktisk stadigvæk at gøre det andet. Fordi vi mener vi
298 bestemt har en bedre case, men så er det en informeret beslutning vi træffer, så vi ved når vi så lancere at så kan
299 der være et problem her, fordi folk har altså gjort opmærksom på det. Jeg kan i hvert fald mærke, når det er den
300 approach vi tager - og det er meget meget nyt, så vi har ikke noget ude, hvor vi har gjort det sådan endnu. Så
301 jeg kan sagtens sidde og spille smart omkring det, men jeg kan også høre det gør også designerne mere trygge,
302 fordi de stadigvæk har et mandat, fordi jeg har lavet rigtig meget på at vi skal satse på at være så brugercentriske
303 som muligt også står der nogle koncept-managers og nogle designere på den anden side også siger "Nej, nej, nej
304 det skal være design-driven det hele" også siger vi jamen okay, men hvor kan vi så mødes? Jamen vi kan godt
305 mødes halvt her, altså næsten på midten, hvor vi så siger, det er fint vi tester det, og det er fint at vi får nogle
306 rigtig gode findings på alle tingene, men det er ikke sikkert at vi hører efter hvad der er blevet sagt. Men det er
307 mere informeret, ikke?

308 **F:** Ja?

309 **P:** Efter jeg har lanceret det, så er det ikke så farligt mere. Fordi det er fint nok, men de har stadig noget at sige,

- 310 ud fra deres faglighed og deres stolthed omkring design.
- 311 **F:** Ja, man fjerner ikke deres behov for at være med indover.
- 312 **P:** Nej præcis, og det begynder lige så stille at løsne op omkring det nu og det er meget mere det her med at vi
- 313 lige går ned og kigger på prototyperne, lige snakker om det og har folk med dernede og sådan, ikke? Men det er
- 314 noget der tager ekstremt lang tid, også det her netop med at gå og vise tingene og få noget feedback, for der er
- 315 meget af det her med "Nej det du'r ikke", fordi folk kan ikke finde ud af at forholde sig til det, hvis det ikke er en
- 316 pæn prototype - og det er jo sådan lidt, det kommer jo an på hvad vi snakker om, fordi hvis det er interaktion, så
- 317 gør det jo ligemeget, altså det der vi spørger hvor mange et eller andet vil I have på noget, for det kan folk godt
- 318 forholde sig til. Noget andet er at forholde sig til om de vil have det specifikke produkt ind og stå i deres hjem,
- 319 fordi hvordan det så helt ser ud, det er helt industrielt design-agtigt.
- 320 **F:** Jaja, lige nøjagtigt, det kommer jo an på hvordan man præsentere det og hvordan man stiller sin test op
- 321 **P:** Lige præcis, det er igen den der.. Tilbage til den her matrix omkring, hvordan er det så man gør det her. Så
- 322 sige, nu er det interaktions vi snakker om, så kan prototypen være sådan her, nu er det selve designet vi snakker,
- 323 jamen så skal fidelity se sådan her ud, i forhold til nogle specifikke faktorer. Også kan det mappes ud på samme
- 324 måde.
- 325 **F:** Nu nævnte du den her matrix igen, tænker du at man måske... den vil jo nok være 2D, men tænker du 3D vil
- 326 være en mulighed i forhold til at man måske opstiller sådan en matrice, også alt efter hvilken side du kigger på
- 327 den fra, så tager den ligesom [...]
- 328 **P:** [Det kunne den sagtens, for jeg tænker virkelig.. Hvis nu vi taler ud fra participants, så har vi noget med
- 329 hvilke krav er der fra... Eller hvilke requirements er der fra kundesiden og hvordan tester vi. Så må der også
- 330 være noget omkring hvilken fidelity der så også skal være. Jeg tænker også tid er et aspekt i forhold til hvor
- 331 modent et koncept er, men jeg tænker mere eller mindre det ligger i fidelity.
- 332 **F:** Ja, og det ligger nok også i valg af metode i forhold til testen, en gang i mellem også - så det er måske implicit.
- 333 **P:** Ja, men spørgsmålet er jo faktisk så, hvis vi tager det der træ, kunne man lave den sådan en 4'er, hvor man
- 334 så siger, vi har nogle requirements, vi har... en metode... metode er der ja, også har vi fidelity der... for så har
- 335 vi, hvis vi nu har metode inden også ikke, fordi så kan vi jo se, hvis vi så vender den på... så siger vi at metode
- 336 - vi kan lige som sige at vi har den her metode og vi ved at det er det her vi vil, vi ved at det eksempelvis er en
- 337 fokusgruppe vi vil holde, fordi det har vi mulighed for lige nu. Hvilk type requirements er det så vi vil kunne
- 338 svare på, i forhold til det og hvilken type fidelity er det så vi skal bruge til det i forhold til. Vi kan også have nogle
- 339 requirements der siger vi skal teste med folk, der er vant til at stikke sig selv. Der er et eller andet her, der godt
- 340 kan blive til noget, tror jeg. Den skal lige gennemtænkes.
- 341 **F:** Ja, det er lige det. Det er dog ikke sikkert vi finder frem til en løsning i dag.
- 342 **P:** Nej, men jeg tror faktisk der er et eller andet - det er lige med at finde ud af hvilke værdier, der skal være på
- 343 de forskellige sider og hvordan de spiller sammen. Jeg tror helt sikkert der er et eller andet - jeg bliver ved med
- 344 at tænke en kube også, men jeg ved ikke lige helt hvorfor.
- 345 **F:** Det kunne sagtens være det, der kunne være løsningsforslaget. Lige kort, bare for at vende tilbage til B&O
- 346 og at de nu lidt skifter side. I prøver vel lidt og løse nogle af de problemer, der har været ved at have et skjult
- 347 design? Jeg kunne godt tænke mig og høre hvordan omgik i dem førhen?
- 348 **P:** Førhen der havde vi nogle koncept-lead og nogle designere, der troede de var guder og vidste hvad folk vil

349 have. Et eller andet sted tror jeg vi har været heldige, at vi nogle gange har haft folk der har haft flair for det
350 lavede, og flair for kunder og flair for usability og UX. Så har vi haft nogle gange, hvor det ikke har været
351 tilfældet. Så det har virkelig været sådan en rutsjebane. Men.. meget stærke personligheder, der har fået nogle
352 fede ting trumfet i gennem, men det har ikke været med fokus på usability og UX.

353 **F:** Nej, på brugerdelen?

354 **P:** Nej, slet ikke og det er netop det vi er ved at gøre op med nu. Vi bliver nød til lige at høre efter hvad markedet -
355 eller om end ikke andet have en idé om hvad markedet mener om de ting vi laver. Men det er jo mange forskellige
356 ting. Der er mange forskellige initiativer, der er on-going lige nu og der bliver rykket rundt hele tiden og der er
357 ændringer og der er organisatoriske ændringer og der er ændringer for hvordan vi arbejder. Vi er i gennem en
358 agil transformation samtidig med - altså der er rigtig meget der rykker og nogle gange så tænker man, hvis man
359 kan holde boldene i luften, så bliver det virkelig fantastisk, men nogle gange tænker man også ej ej, det er alt for
360 meget på en gang, det hele må falde til jorden på et eller andet tidspunkt. Der sker rigtig mange ting, synes jeg og
361 netop især på området med brugerinddragelse og... vi snakker rigtig meget people before technology, fordi det
362 har vi været slemt til lige som at sige at det var folk eller vores kunder skulle egentlig bare indrette sig efter den
363 teknologi der nu en gang var og det design vi havde valgt for dem. Vi bliver nød til at vende den her tankegang
364 om, også sige hvem er de folk vi udvikler til og hvad er det egentlig de har behov for.

365 **F:** så lidt over i samme boldgade igen, altså med de her identity-roles?

366 **P:** Øhm, et eller andet sted, ja lidt. Vi er i hvert fald i gang med meget at få mappet ud, hvilken kundegrupper det
367 er vi primært designer til også sørge for ligesom at få målrettet vores produkter i forhold til at nu er det det her
368 segment, hvad er det så der er vigtigt for det segment? Så tester man med det segment, men stadigvæk har også
369 øjnene åbne for at det også er andre end det segment der skal købe vores produkter. så der er mange forskellige
370 ting, også er det hele det der med segmenter, er der forskel på om man snakker marketingssegmenter kontra de
371 segmenter vi udvikler til? Det burde der ikke være, men det er der tit alligevel. Vi vil jo gerne have noget der
372 er mere persona-rettet hvor de jo har noget der er mere arketyptisk rettet. Så der er mange ting der er lidt det
373 samme også alligevel ikke helt.

374 **F:** Ja okay

375 **P:** Men mange forskellige ting. Jeg... Agile UX er jo det jeg kan, og det jeg prøver at få indover - så det er meget
376 med at få nogle processor omkring det. Det har været meget ad hoc det arbejdet der er blevet gjort, så at få nogle
377 processor op at køre sådan generelt, nogle processer om at få ting sat i en agil kontekts også.

378 **F:** altså det lyder super spændende og jeg vil også sige det er meget brugbart i forhold til, hvad jeg arbejder med.
379 Men i forhold til det her med at der var nogle designere og nogle ledere, der følte sig som guder og måske havde
380 lidt for meget ejerskabsfølelse, var der så nogen som helst struktur der, i forhold til at det var ekspertbedøm-
381 melse? Var der metoder på eller var det bare gut-feel?

382 **P:** Altså der har været noget - men jeg tror også det har været meget person-bestemt og meget ad hoc også har
383 det også været meget med 'måske bliver der hørt efter, måske gør der ikke - vi ved det ikke helt'. Altså jeg kan
384 sådan se, jeg har været her lidt over et år nu, og vi har nogle produkter, hvor at de ligesom har rejst et skrig, hvor
385 jeg sådan har sagt, at der er et eller andet, som ikke er helst som det skal være her også finder jeg ud af at der
386 faktisk har været lavet brugerundersøgelse omkring det, der er bare ikke blevet hørt efter de findings, der så kom
387 ud af det. Så der har været lavet, men det er som om der ligesom er blevet sagt "Ah, hvad ved de? Ah, vi gør bare

388 som vi plejer".

389 **F:** Det er jo også en genial fremgangsmåde jo.

390 **P:** Amen, lige præcis ikke? Der har været lavet noget og jeg ved os når jeg snakker med de UX'er vi har, der har
391 været mange initiativer, som er startet op også måske er dødt ud lidt igen også er startet lidt op igen. Så det er
392 lidt, da jeg blev ansat, nu kommer der endnu en, der prøver at gøre noget. Så nu håber jeg bare at jeg kan holde
393 den gående et stykke tid.

394 **F:** Ja

395 **P:** Men nu er tiden også ved at være mere moden til det. Da jeg lavede min P.H.D og lavede nogle interviews
396 med forskellige virksomheder, der er også den der med at det kan godt være at folk eller virksomheder siger, at
397 vi skal have fokus på usability og UX, men det er faktisk først når man virkelig har haft et produkt der er fejlet
398 bigtime, hvor man så får øjnene op for at det kunne vi faktisk godt have fanget, hvis vi havde gjort noget i den
399 retning. Det tror jeg måske, ikke fordi vi har haft nogen produkter der har været sådan, men nu er tiden også
400 ved at være moden til, at virksomheder og forretninger kan se at de bliver rent faktisk nød til at inkluderer en
401 eller anden form for slutbruger i det arbejde der bliver lavet.

402 **F:** Er det udelukkende set fra B&O's synspunkt eller synes du det er sådan mere generelt?

403 **P:** Det er generelt. Jeg synes generelt usability user maturity er... Jeg havde faktisk et møde i går, også med en,
404 hvor vi snakkede om det - der har været meget snak om det, men ikke så meget gjort omkring det, hvor jeg tror at
405 nu bliver det stille og roligt også en del af nogle processor, fordi det er sådan vi vil gerne og alle kan se idéen i det,
406 men så når man har travlt og forskellige ting, så er det lidt det der bliver udladt og så er det sådan lidt på gefühl
407 og adhoc også gør vi det lidt hist og her. Jeg synes at der er lidt en tendens til nu at der begynder at komme lidt
408 mere process op omkring; hvornår gør vi det så og hvordan gør vi? Hvor sådan et tool - hvis du fik lavet det her,
409 det ville også været helt ideelt. Hvis man så kunne adoptere det til andre virksomheder også. Og hvis man bare
410 kunne sige vi har den her kube eller det her sheet - også kan man egentlig bare selv begynde at putte noget ind
411 i den, ikke? Helt ideelt, også fordi så bliver det lige pludselig ikke så farligt. Jeg kan mærke her, at det er rigtig
412 meget rekrutteringen, der virkelig er et stort prblem ofte, ikke? Jeg arbejde ved Radiometer før jeg kom her, hvor
413 det også er medical devices lidt lige som Novo - ikke helt som Novo, for det er mere medicinsk - men også hvor
414 det var meget snævre target groups, hvor det var svært at finde testpersoner, man bookede gennem bureauer, det
415 tog en krig og det kostede en bondegård osv.. Hvor jeg egentlig havde troet at det ved B&O ville være let, fordi vi
416 har kunder. så var der en business-model der sagde at man ikke måtte snakke direkte med kunder og dit og dat,
417 så den har vi så også fået gjort op med nu. Men stadig den der med at det tit er rekrutteringen der er hæmskoen
418 og jeg kan ikke helt lure om det er fordi det er det der er lettest at sige eller om det faktisk reelt er dét det er. Jeg
419 ser den virkelig som en hæmsko.

420 **F:** Hvis man bare lige tager rekrutteringen sådan ganske kort - man kan sige selvfølgelig er det svært at rekrut-
421 tere, hvis man ikke ved hvem man skal rekruttere efter, men ser du andre problemer i det? For det er jo i realiteten
422 bare at tage kontakt.

423 **P:** Ja, men det er så det der med at der er nogle specifikke kriterier og det er et specifikt segment du gerne vil
424 teste med, ikke? Så har vi den problemstilling, men så kan man så sige, hvad så hvis vi får det afdækket, hvad er
425 det så der gør det? Det er så her kæden den hopper lidt af, fordi der burde virkelig ikke være noget der ligger til
426 grund for det. Også her ved B&O, vi kunne jo reelt set bare tage folk ind fra gaden - eller til noget af det ville vi

427 kunne.

428 F: Ja, lige nøjagtig.

429 P: Det underer mig bare, at det ikke er det der sker. Der er en barrierer af en eller anden art, hvor det skal gøres let tilgængeligt. Jeg har været i gang med, nu er det så på hold lige nu, men netop med at lave sådan en participant pool, hvor det simpelthen vil være let og så ligesom kunne sige at det er folk der er interesserede i at deltage og som har signet sig op også vil man kunne skrive en eller anden fællesmail ud; "Hey vi tester i Struer mandag, tirsdag, onsdag - her er kalenderen, i kan bare booke jer selv hvor det passer". Det kan jeg mærke at det gør det lidt lettere, vil jeg sige. Men generelt er rekrutteringen virkelig en hæmsko i det her arbejde og der er også et eller andet - jeg ved ikke om det er noget med at folk er lidt generte og det der med lige at få spurgt. Generelt er folk - også internt - villige til at sige "okay, jamen jeg kan da godt lige komme og en test her og nu". Det er ikke alle steder det er sådan, der hvor jeg arbejde før, der var det sådan "Nej, det har jeg sku da ikke tid til, hvad fanden tror du?". Så der er et eller andet der, men jeg er ikke helt sikker på endnu, hvad det er - men jeg synes at jeg begynder at kunne se lidt en tendens deromkring og at det tit er der hæmskoen er.

440 F: Ja, bare det der med simpelthen at få fat og tage kontakt?

441 P: Ja, og der er mange aspekter i det, men den helt lavpraktiske er at man er bange for at man tager folks tid, har folk tid, folk har også travlt og dit og dat - det er en del af det. Der er bare også noget andet i det end det - jeg kan ikke helt pinpointe det endnu, men jeg kan begynde at.. De snakke jeg har med folk, hvor jeg kan høre det - når nævner de ting, så de statements der kommer de supporterer ligesom det der - så der er et eller andet.

445 F: Ja, så der begynder at komme lidt på hvorfor det egentlig...

446 P: Ja, lige præcis og jeg kan også høre at det jo er mange steder det er sådan og det er virkelig underligt hvorfor! Så snakkede jeg også med nogle i går netop med - jamen er det også fordi at vi vil gøre det så godt og vi er bange for at fejle hvis vi rekrutterer interne folk? Fordi jeg tror også der er meget den der med at vi UX'ere er lidt nogle primadonnaer en gang i mellem. Vi har lært at man gør det på den her måde, så det er ved gud at vi gør det på den her måde. Så hvis vi skal til at begynde at slække på noget... nogle gange tænker jeg også det er bedre at have én end ingen og teste med og det er også nogle gange bedre at ligesom at sige, så tester jeg den her lille del af det. Det kan godt være det ikke er det hele jeg tester, men det er det er den her lille del også ved jeg i hvert fald at det her - det virker. Så tror jeg også nogle gange at vi har en tendens til at sige, hvis vi ikke kan teste det hele, så kan det også bare være ligemeget, ikke?

455 F: Jo og ligenøjagtig. Jeg vil jo også gerne prøve at undgå den der lidt primadonna-agtige følelse af at det skal være den rigtige, ellers er der ingenting der dur'

457 P: Jamen netop. Den hører jeg rigtig meget og jeg synes især også - når jeg har arbejdet meget med agilt UX, de akademiske metoder der er, i forhold til at lave usability og UX tests, de er meget meget tunge og de er meget akademiske, det er noget der tager lang tid og det er noget der kræver relativt mange ressourcer også. Så det der med at få metoderne lavet lidt mere agile og lidt mere lette, det er altså også vigtigt. Det var det jeg arbejdede på med min egen P.H.D. Hvordan kan man modifcere de her metoder - det kan godt være det ikke er validt fra et akademisk perspektiv, men det er i hvert fald nogle findings der er valide nok til at kunne træffe nogle beslutninger ud fra.

464 F: Ja, lige nøjagtig. Så længe man bare har lidt information at træffe sine beslutninger ud fra og at man ikke bare gør det på gut-feel.

- 466 P: Lige præcis. Det arbejdede jeg jo så med i 3 år og jeg underviste også i det, hvor jeg kunne høre at det er
467 noget der resonere rigtig meget med de virksomheder, der sidder og arbejder med det. Det er den der med... Så
468 tager man den der dybe indånding og tænker "så skal det også være det der stooore user study" i stedet for det
469 bare... Det behøver det jo ikke, det kan jo bare være 5 minutter og en håndtegning, hvor vi lige gennemgår noget
470 og hvorfor. Så kan man sige "Ahhh det der giver simpelthen ikke mening, fordi det er ikke sådan man normalt
471 ville gøre det og det her er altså indutristandarden indenfor... det er jeg da vant til" - også ser det som en finding
472 og at det ikke bare er ligegyldig information, for det er det jo ikke. Og også italesætter det som at det faktisk er
473 en eller anden form for brugerstudie.
- 474 F: Jo, at man ligesom i bredere grad acceptere at brugerstudier kan gradbøjes og det ikke er...
- 475 P: Lige præcis! Der tror jeg også der er et paradigmeskifte - jeg tror det er i gang, men jeg tror ikke det er
476 helt fuldendt endnu. Det er ligesom at man kommer fra 'Det er sådan her man gør, eller også gør man det ikke
477 ordentligt også gør vi det ikke' også til 'Vi kan også bare gøre sådan her'. Lidt ligesom det her med netop den
478 her kuge eller hvad det bliver til. Hvor man kan sige det kan godt være det ikke er det hele der er testet, men
479 det kan være det er 75% der er testet også har vi så en marginen på 25% der ikke er testet, men hvor vi enten
480 konkluderer at det burde ikke være noget problem eller vi ved der kunne ske at hænde fejl, når den ender ved
481 den rigtige slutbruger. Men netop det der med at kunne differentiere i det; det er ikke 100% men så er det 75%
482 eller hvad det nu kunne være. Hvis man kan teste 75% med folk der ikke er i target gruppen også de sidste 25%,
483 hvis de skal testes, så er det der man bruger de folk der kan bruges til det. Et eller andet sted, så er det jo sund
484 fornuft, ikke?
- 485 F: Jo, lige nøjagtig, men jeg tror også lige så meget at det er at få overbevist folk om at man godt kan bruge den
486 her sunde fornuft og at det ikke altid skal være som det står skrevet op og ned.
- 487 P: Lige præcis. Det er virkelig også... Jeg prædiker meget om det og det resonere også fint, men der er stadig
488 lidt en barrierer omkring de ting der - er det så validt nok og tør jeg med den baggrund eller position at stå inde
489 for at det er validt nok, når jeg ikke har gjort det 100% som jeg er blevet lært?
- 490 F: Ja, men et argument kunne jo være, at nu er beslutninger taget på noget og ikke bare på ingenting.
- 491 P: Lige præcis og det er netop det der er vigtigt, ikke? Men jeg synes virkelig det er spændende!
- 492 F: Det er jeg glad for at høre
- 493 P: Ja, og det kunne være være et åndssvagt fedt tool at få, det her, altså.
- 494 F: Ja, jeg glæder mig jo til at se, hvor det bærer hen af. Man kan jo sige grunden til at jeg lige som har taget fat i
495 dig, har jo været at prøve og få lidt empiri på, hvad der egentlig er problemet, fordi litteraturen ikke understøtter
496 det. Så det har givet meget!
- 497 P: Skidegodt! Og det er jeg glad for, selvom det blev sådan lidt en rambling over alt muligt forskelligt!
- 498 F: Det tror jeg det er det der skal til, for at komme fremad, for lige nu er det bare, hvis du ikke har den rigtige
499 brugergruppe, så kan du lige så godt lade være.
- 500 P: Ja, og det er jo fuldstændig skørt ikke?

Transcript of Interview with Morten Purup

- 1 **F:** Hej Morten,
- 2 **P:** Hej Simon,
- 3 **F:** Jeg skrev til dig, for at snakke om det her sampling, og man kan sige baggrunden for det er, at jeg lige har
- 4 afsluttet et halvt års praktik i Novo Nordisk, hvor jeg ligesom har set nogle problemer i rekruttering og sampling
- 5 og steder, hvor det synes nærmest uladsiggørligt.
- 6 **P:** Ja...
- 7 **F:** Så jeg er ved at skaffe lidt empiri på hvad der egentlig rør sig i andre virksomheder og snakke med folk der
- 8 ligesom har arbejdet med det. Man kan sige at det problem jeg så til at starte med, er at der bliver kastet rigtig
- 9 mange ressourcer efter alt det her rekruttering og at man skal nå de rigtige brugere. Og det næste er så det her
- 10 med at når man designer til en population, som stort set ikke kan nås - som måske gør sig gældende ved sjældne
- 11 sygdomme - hvordan får man så ligesom testet det man er i gang med at designe? og det er nok mest det device-
- 12 baserede jeg kigger på.
- 13 **P:** Ja...
- 14 **F:** Så det er ligesom baggrunden for det
- 15 **P:** Okay, spændende
- 16 **F:** Det jeg så har arbejdet hen imod er at hvis man har en eller anden type evaluering - lad os sige field evaluation
- 17 - og man ligesom har sin test klar, men at man ikke kan nå den her population, hvad der så kan gøres af tiltag for
- 18 at få nogle resultater, som stadig er valide. Så til at starte vil jeg gerne høre hvilke problemer der gør sig gældende
- 19 i forbindelse med det du arbejder med. Jeg ved du har været ved Radiometer tidligere...
- 20 **P:** Ja... Jamen det har jo også været ret snævre brugergrupper vi har skulle rekruttere nogle gange, hvor man
- 21 også er rendt ind i problemer, som følge af ikke at have brugt de rigtige testdeltagere. Så derfor blev der, i hvert
- 22 fald til sidst i den tid jeg var hos Radiometer, ikke stillet så meget spørgsmål til rekrutteringsudgifterne til at få
- 23 de rigtige testdeltagere, altså nogle der har en hvis erfaring med en bestemt type apparatur eller ja... specifikke
- 24 behov osv. Det er simpelthen fordi man har kunne se de tidlige problemer man har haft ved og rekruttere
- 25 for brede brugerprofiler fremfor nogen... Simpelthen fordi der har været og er for store forskelle. Hvis du bare
- 26 rekruttere sygeplejersker der er van til at håndtere blodprøver - der er kæmpe forskel på om det er hæmatologi
- 27 eller om det er blodgas analyser for eksempel. Vi har bare fundet ud af at vi simpelthen bare ikke kunne bruge
- 28 resultater fra for eksempel hæmatologi-sygeplejersker og de har ikke været repræsentative for de intenderede
- 29 brugere som vi skulle ramme. Det har været et problem i forhold til egentlig at kunne konkludere - kunne clame
- 30 - at vi havde et sikkert og effektivt apparat betingelsesmæssigt overfor myndighederne.
- 31 **F:** Ja, hvad med... Nu siger du myndighederne, det lyder meget som en summativ evaluering - hvad med de
- 32 mere formative studier, hvordan har det set ud der, i forhold til sampling?
- 33 **P:** Jamen det har været alt fra at vi tester på to ved nabo-skrivebordet eller nogen man finder i en anden afdeling, finans eller HR afdelingen som intet ved om, hvad der foregår i udviklingsafdelingen. Så det kan være dem,
- 34 det kan også være... På radiometer havde vi laboranter internt, så det er jo egentlig super-brugere som kan
- 35 repræsentere. Nogle gange ved de jo for meget, men... Ja det kommer jo an på hvad man tester, ikke? Fordi hvis
- 36 du tester noget med om det er en stor nok skriftstørrelse på vores GUI, jamen så kan du tage Fru Jensen fordi de
- 37 perceptuelle egenskaber og de forudsætninger hun har for at læse, er ikke anderledes end andre sygeplejersker,
- 38 da det er basalt ens for de brugergrupper. Men hvis det er noget, der typisk har med det kognitive at gøre, noget
- 39 man skal vide i forvejen eller noget man skal være særligt obs på eller noget der har med specifikke regler eller
- 40 problemløsning eller omsættende information at gøre, så gælder det simpelthen om at få nogen der ved hvad de
- 41 har med at gøre.
- 42 **F:** Ja, altså jeg vil heller ikke foreslå at ligemeget og ligegyldigt hvad man evaluerer og hvilken metode man
- 43 bruger så kan man bare erstatte, fordi nogle gange der skal have de rigtige brugere - det er jeg helt med på. Det

45 er mere den her tilnærmelse, altså så snart det er value-based, hvad det kan give af værdi og sådan for brugeren -
46 lad os sige et meget specifikt brugsscenario, så er det selvfølgelig klart at den rigtige bruger skal inddrages, men
47 jeg tænker at det er lidt på en skala og alt efter, hvor langt man er i processen, så graduerer det lidt i hvilken
48 sample man egentlig kan tillade sig at benytte?

49 **P:** Jeg vil mere sige det kommer an på hvad du undersøger, fordi det der med hvor langt du er i processen det er...
50 Der kan være behov for at ramme lige præcis den intenderede brugergruppe tidligst i processen, men også... Ja,
51 så tidligt i processen, at du stadig kan lave noget om. Jeg tænker eksempelvis i forhold til user needs og identifi-
52 cation, der er det vigtigt at have de rigtige. Så kan der være noget i de aller tidligste faser, hvor man lægger sig
53 fast på noget koncept, hvor det også kan være problematisk, hvis ikke man har den rigtige brugertype.

54 **F:** Ja, man kan sige, der hvor jeg står lige nu, altså det er lidt det her med at, - som du siger - at der er steder,
55 hvor man skal have den rigtige brugertype. Også kigger jeg lidt ind i det der med... Nu sagde du selv at hun er
56 ligesom enhver anden bruger og havde der perceptuelle egenskaber der var nødvendigt for at udføre... Det her
57 med om man måske kan analyserer den population man kigger på, få identificeret de her karakteristikker for den
58 sygdom eller for den person også lede lidt bredere end lige den specifikke sygdom eller person - så længe at de
59 her perceptuelle og kognitive egenskaber er nogenlunde opfyldt.

60 **P:** Ja... Jamen selvfølgelig er der områder, hvor du kan bruge andre personer og det... Så er der det her eksempel
61 med legibility eller om fontstørrelsen er stor nok eller lysforhold og sådan noget. Der kan man bruge andre. I
62 forhold til, måske, præference af noget look-and-feel kan det også være man kan benytte nogle andre.. i forhold
63 til navigation i en software-GUI kunne det også være man kunne benytte nogle andre, og sådan noget som...
64 Altså sådan noget som () og alt det antropometriske.

65 **F:** Ja - I forhold til de problemstillinger identificeret ved at bruge for brede samples, hvad var de typiske prob-
66 lemstillinger så?

67 **P:** Jamen dels var det at de simpelthen ikke kendte til hvilke faldgruber der er ved en blodgasanalyse. Der er det
68 man - det FDA kalder knowledge tasks - der var simpelthen nogle steps inden man analyserer blodprøven, som
69 hvis du ikke udfører dem eller udfører dem i en forkert sekvens så udgør de en risiko. Der var også udfordringer i
70 forhold til at protokollen, testprotokollen, havde statet specifikke inklusionskriterier til testdeltagerne, som ikke
71 blev efterlevet og det var sådan noget som kvalitetssikring og (registrerings...) opponerede meget imod.

72 **F:** Ja. De manglede jo så lidt den her læring og kontekts omkring brugen af jeres apparat. Tænker du det er noget
73 man måske kunne have frontloadet i forhold til og gennemfører en test og få valide resultater derigennem?

74 **P:** Ja, altså alternativet er at du skulle oplære naive testdeltagere til at have samme viden, som repræsentative
75 brugere - det tror jeg bare ikke er realistisk, ikke i det her tilfælde i hvert fald.

76 **F:** Nej, altså man kan sige, sådan noget som panel-testing og sådan noget, det er jo lidt med oplæring og sådan
77 noget.

78 **P:** Panel-testing?

79 **F:** Ja, det er hvor man samler en masse brugere - eller folk melder sig til og gerne vil teste også får alt efter hvad
80 det er, en eller anden oplæring inden for emnet. Men så vidt jeg har læst mig til, så har man det her panel også
81 med tiden, så bliver de jo også eksperter i det - Så det er jeg lidt gået væk fra igen og kigge nærmere ned i, for det
82 er det her med at have et kæmpe panel og sikre at man ikke fik de samme - så man lige pludselig havde de her
83 eksperter. Men meget apropos eksperter, hvordan sådan... Der er mange der ligesom sidestiller field evaluation
84 og laboratory evaluation, hvor du bruger rigtige brugere, med de her ekspert brugere fra eksempelvis cognitive
85 walk-through og heuristic evaluation. Hvordan forholdet der sig ved jer? Er det noget i har kigget ind i, eller?

86 **P:** Øh ja. Den metode vi har fået mest ud af er egentlig task-analysen. I bund og grund kan du sige, der er ikke
87 nogen grund til at gå ud og teste ting, som en mand kan finde ud af bag skrivebordet. Så vi benytter de analytiske
88 metoder, som brainstorm og task-analyse som... altså vi benytter dem på den ene side også de empiriske, som
89 usability test og det at indsamle data fra rigtige brugere til at kompensere hinanden. Til at nå begge veje rundt,
90 ikke?

91 F: Ja. Hvad siger du, task-analysen er det simpelthen hvor du går ind og kigger på den enkelte opgave eller
92 hvordan forholder det sig?

93 P: Ja, altså du tager hver task og bryder ned i diskrete steps, for eksempel hvis det er et eller andet apparat så
94 kunne det være noget med at tænde det, eller at klargøre det også kunne det være noget med måle også kunne
95 det være noget med at få resultatet videre eller ud eller sendt til et IT-system. Det kunne også være for en for-
96 brugsvarer, du skal finde den her forbrugsvarer, så skal du aktivere den også skal du installere den, Lidt ligesom
97 en printerpatron eller sådan et eller andet.

98 F: Ja og der får i så bare en... hvad skal man sige, en ekspert til at???

99 P: Ja, altså det kan laves af en person også kan det selvfølgelig reviewes af andre. Så er det jo for så vidt muligt
100 også at få tænkt konteksten ind i det, lysforhold, lydforhold, afbrydelser... Hvis det er i en ambulance, så kan
101 det være de kun har én hånd til at interagere med devicet med, fordi den anden hånd brugere de til at holde fast
102 med, mens der køres stærkt.

103 F: Ja, så det vil være lidt det der, hvis du ikke tog højde for det, i dit laboratorie-setup, så ville du ikke få valide
104 resultater, eksempelvis?

105 P: Ja for eksempel. Task-analysen er så den analytiske metode, så man kan identificerer noget af det her, inden
106 man går ud og laver usability-tests.

107 F: Ja - i forhold til det her med konteksts, så synes jeg - nu har jeg kigget en del studier i gennem, og der er blandt
108 andet nogen der sammenligner field evaluations og laboratory settings og det her og jeg synes det de kommer
109 frem til er, at ca. halvdelen plus det løse kan identificeres i gennem de forskellige ting, men så er der en spredning
110 på, hvad der bliver identificeret gennem de forskellige metoder. Eksempelvis field er der meget konteksts og lab-
111 oratory, der er det nogle andre forhold der gør sig gældende - måske det at de sidder i en stol og gennemgår det.
112 Tænker du at der er nogen work-around i forhold til det - altså... Mange af de samme usability problems bliver
113 egentlig identificeret gennem de forskellige metoder for eksempel field, lab og heuristics og efter ca. halvdelen,
114 så begynder det at være lidt mere spredt ud - så er det nogle forskellige problemer de får øje på. Tænker du at der
115 er en work-around der kan gøre op for det eller tænker du at metoderne ligesom skal komplimentere hinanden
116 i forhold til det?

117 P: Jeg tænker at de komplimentere hinanden, for der er nogle ting i en usability test, som du bare ikke kan
118 tænke dig frem til. Vi havde for eksempel... jeg kan lige komme med et eksempel først, også kan jeg komme
119 med hvordan man måske kan komme omkring det analytisk. Vi havde et eksempel hvor.. for at måle på en
120 blodprøve, så skal man trykke sprøjten hen på et inlet - altså væk fra sig selv, man skal simpelthen trykke inlettet
121 op. Der sidder sådan en lille fjeder der aktiviserer opsuget. Den her action eller den her sekvens, den er også
122 illustreret med en video, hvor man ser en videosekvens af hånden der gør det her. Til at starte med så den her
123 video-sekvens den loopede, så vi så simpelthen i brugertesten at folk de stod og trak hånden frem og tilbage, fordi
124 videonj loopede. Så stoppede vi så videon, efter den var færdig og gjorde den grå og satte sådan et replay-ikon
125 på. Det replay-ikon, det var på det tidspunkt hvor YouTube havde sådan en cirkel - sådan en pil rundt i en cirkel.
126 Vi så så en testdeltager stå foroverbøjet over det her apparat og forsøgte at skrue sprøjten på - simpelthen som
127 om der var et gevind. Vi spurgte så, hvad tænker du nu, eller hvad prøver du på? siger så "Den siger jeg skal
128 skrue den på". Det havde vi jo slet ikke forestillet os gennem brainstorm og task-analysen. Der hvor man måske
129 ville kunne identificerer det i de analytiske metoder, det vil være ved at kigge på hvilke mentale modeller har
130 folk og hvilke mentale modeller udstråler designet, eller ikoner osv.

131 F: Så hvad skal man sige, lave ligesom et bredt sprektre af hvilke mentale modeller der kunne gøre sig gældende
132 når man bliver præsenteret for det og samtidig lave en analyse af selve udseendet af produktet? Ja det hænger jo
133 nok lidt sammen

134 P: Ja, opg hvilke mentale modeller det vil kunne forandre, ja.

135 F: Det jeg tænker er, at du også kommer ud i de situationer - i hvert fald med noget af det medicin der laves
136 på Novo Nordisk og de devices dertil, er at der ligesom ikke er koncentreret mængder af brugere nok til at man

137 rejser til én stat, tester også rejser hjem igen, men det her med at du faktisk skal rundt til de forskellige nærmest.
138 Om man kan komme rundt om det på en eller anden mere fornuftig måde; ét i forhold til ressourcer, men også
139 to, det her med du siger at konteksts aldrig vil være holdt konstant, hvis man ligesom rejser rundt til forskellige
140 dele og nye usability forhold osv.

141 **P:** Jeg tænker ikke det er forholdene der er udfordringen der, det er mere det med at rekruttere de rigtige. P.t.
142 er jeg ved at planlægge et usability studie for en kunde, hvor vi regner med at skulle rundt til... vi regner med
143 at kunne rekruttere fem til seks deltagere i hver større by, så vi skal rundt til en... Jeg tror vi har regnet med fem
144 større byer i USA, til den summative test. Det er simpelthen fordi de er svære at komme i nærheden af, de har
145 super svagt immunforsvar og det er generelt svært at få dem ud i testlokaler og der er også risiko for at de smitter
146 hinanden og så videre.

147 **F:** Ja, men hvilke forbehold tager I jer så, i forhold til det? Altså simpelthen det her med, ét det her med at de
148 ikke kan komme ud, men to, også at der er ret mange variable der ændrer sig, hvis I skal ud.

149 **P:** Jamen altså, nogen af dem kan jo ikke komme ud og det er dem vi gerne vil rekruttere og vi ved også at det
150 rekrutteringsbureau vi har fat i, har lavet studier med den her brugertype før. Men ellers siger de at det typisk
151 er telefoninterviews de laver med den her brugergruppe, men nu har vi altså et device med et nyt koncept som
152 skal testes også er der ikke rigtig nogen vej udenom, i forhold til at efterleve kravene fra myndighederne.

153 **F:** Tager I jer så nogle forbehold i forhold til testen? Prøver I ligesom og sige at i kender... Der vil være de her
154 begrænsninger i forhold til det der og ligesom prøve at godtgøre på det forhånd eller?

155 **P:** Ja øh... Altså viu har taget de forbehold at budgettet er større og har sat flere penge af til at rekruttere og hvis
156 vi skal bruge altså pr. 15 testdeltager skal vi bruge 18 i tilfælde af aflysninger eller folk der ikke dukker op. Vi har
157 sat både tid og penge af til og rejse til seks større byer i stedet for én eller to. Så hvis man skal tage uyderligere
158 forbehold, så skulle man sætte endnu flere penge og mere tid af og besøge flere byer ud fra en worst-case be-
159 trætning af at man måske kun får to testdeltagere i hver by.

160 **F:** Ja, det vil ikke være så eftertragtet kan man sige. Men jeg tænker også det her i forhold til med at nu rejser
161 i rundt til dem og omstændighederne er måske anderledes i den ene by i forhold til en anden. Er det noget i
162 ligesom tænker ind i de resultater i nu en gang får?

163 **P:** Det er jo vores opgave og sørge for at konteksten og testmiljøet, testlokalet og hvad der ellers er... Det der
164 er relevant og kan spille ind på interaktionen har meget... At vi kontrollerer dem og simulerer dem som værende
165 ens. Der kan det stille nogle krav til testfaciliteterne og til rekrutteringsbureauet, hvis det er dem der sørger for
166 testlokale. Der er så alligevel lidt begrænsninger, for eksempel åadan noget som gulvtæpper... De fleste af de her
167 typer lokaler, de har desværre gulvtæpper, hvilket gør det sværere at desinficerer sådan et rum.

168 **F:** Ja, og det har så påvirkning på de personer der skal komme, grundet deres immunforsvar?

169 **P:** Ja, det skal gerne være nogle lokaler der er lette at spritte af.

170 **F:** Hvad med i forhold til det her med at lave de her interne tests? Der har i vel også vidst at der var nogle
171 begrænsninger i de data i har fået i forhold til enten har de vidst for meget eller også har der været nogle biases?
172 Jeg ved eksempelvis at et bias i Novo Nordisk i forhold til de interne tests er.. Eller det er i hvert fald et dilemma
173 som er omdiskuteret, det er det her med at hvis man beder en medarbejder fra virksomheden af, om at bruge
174 et nyt koncept, så ligger der ligesom også noget i for dem at "Åh nej, nu skal jeg bruge et produkt fra noget jeg
175 egentlig arbejder for"... Altså ikke frygt, men ligesom at man kommer til at se dum ud. Synes det ude ved jer at
176 være validt nok at lave de her interne tests?

177 **P:** Ja, alt efter hvad vi testede og hvad vi ville undersøge, så ja. Vi benyttede blandt andet HR-medarbejder til
178 eye-tracking studier, hvor vi også havde laboranter som repræsenterede superbrugere. Nu ved jeg ikke om det
179 er sådan et særligt kendetegn for en HR-medarbejder, men de er ikke så bange for at tage ansigt. Det kan der
180 selvfølgelig være andre brugergrupper der er, men man kan sige det kommer også an på om det er større de-
181 signbeslutninger der skal træffes, så basere vi dem nok ikke udelukkende på tests.. altså hallway tests i gangen,
182 der lige er udført. Der har også været nogle gange, hvor måske produkter eller en software chef har, simpelthen

183 krævet at der er lavet brugertest med rigtige brugere, hvis det er at de skal lave en større designændring, som
184 ikke er planlagt.

185 **F:** Okay. Jeg startede sådan med og have en meget teoretisk tilgang til det og læste op på sampling og popula-
186 tioner. Jeg synes ligesom at al litteratur siger det her med at samplen skal være helt repræsentativ, for at det du
187 får ud det reelt set kan generaliseres. Jeg høre lidt på dig som, at det lige som kan efterkommens, hvis konteksts
188 og de perceptuelle og kognitive faktorer de ligesom er vedligholdt.

189 **P:** Njah... Ikke i en validering. Altså i forhold til hvis du er der, hvor du skal finde ud af skal vi have punktstør-
190 relse 10 eller 12 i vores labelling eller sådan noget, jamen så kan du godt hive nogen ind fra nabokontoret.

191 **F:** Okay - så det er meget hvad du tester der afgøre det?

192 **P:** Ja og graden af bevisbryde som testen egentlig skal fremlægge.

193 **F:** Ja, så hvad det skal bruges til bagefter?

194 **P:** Ja, altså om det er en validering eller det er en... noget mindre vigtigt.

195 **F:** En anden problemstilling jeg arbejder ud fra, det er det her med, hvis du har et produkt, der bliver så konfi-
196 dentielt at det ikke kan vises udenfor firma-konteksts. Jeg ved ikke om det er noget der gjorde sig gældende ved
197 Radiometer eller der hvor du sidder nu?

198 **P:** Altså vi har sådan en standard non-disclosure agreement, som testdeltageren skriver under.

199 **F:** Ja og så kan alle deltage når den er skrevet under?

200 **P:** Ja. Altså jeg vil hellere have god feedback på min ide, end jeg vil holde den sådan fuldstændig hemmelig.

201 **F:** Ja, altså idéen fremsprang fra min vejleder fra Aalborg, som er tilknyttet B&O og derhar der ligesom været...
202 Ikke en skjult agenda, men når designet lå sådan nogelunde fast, så blev det meget at nu kan det ikke vises til
203 nogen og nu er det ligesom in-house. Han havde sådan lidt en idé om hvordan det kan løses og hvordan kan det
204 testes og stadig få valide resultater, uden og inddrage både end-users, men egentlig også almindelig users.

205 **P:** Ja... altså jeg synes ikke rigtigt præmissen holder. Hvis testdeltageren underskriver en fortrolighedserk-
206 lærings, så må man sige det er godt nok. Jeg synes lidt det er noget fis at man... når man fastlægger designet så
207 ikke vil vise det til nogen. Nu ved jeg godt der selvfølgelig kan være mange stakeholders, men jeg synes lidt det
208 er noget fis. Det er sku et selvskabt problem.

209 **F:** Nej, og det er heller ikke fordi jeg er uenig, man kan sige jeg interviewede der og de er jo selv ved at ændre
210 tilgang - jo flere øjne på det, jo bedre - om det så er folk der bor i Struer eller om det er end-users, det tager de
211 ikke så tungt mere sådan som jeg forstod på dem, bare der rent faktisk er nogen der gider give input. Det er klart,
212 hvis du også har den holdning, så begynder det da at tale derhen af og det er ligesom også det jeg synes jeg hørte
213 derfra.

214 **F:** Men tænker du i forhold til - hvis vi siger det er et scenerie, at der kan gøres noget i forhold til at du udelukkende
215 ekspert-bedømmer noget eller tænker du simpelthen at end-useren skal med indover eller er det nok bare endelig
216 og validerer? Er der et eller andet man kan gøre for at det vil være okay?

217 **P:** Altså du kan godt... Der er eksempler hvor du godt bare kan evaluerer eller have en ekspert evaluering,
218 det mener jeg bestemt. Der hvor jeg typisk har brugt det, er ved mindre designændringer, hvor du for eksempel
219 kan sige at vi har evaluert brugen af det her produkt og nu ændrer vi kun den her ene del, for eksempel lad os
220 sige bagcoveret skal nu hvor mørkeblåt i stedet for bordeaux rødt, så kan du lave en ekspert evaluering af det,
221 ikke? Men mit fokus har meget været på det her med sikkerheden, altså der er nogen designændringer der let
222 kan argumenteres for ikke at påvirke den måde hvor programmet betjenes på.

223 **F:** Ja, så det bliver meget kosmetiske ændringer?

224 **P:** Ja

225 **F:** Ja, altså man kan sige det jeg arbejder frem imod, er hvordan jeg skal gøre det an, hvis den egentlige popu-
226 lation man designer til, ikke er tilgængelig; enten fordi man er limiteret af ressourcer eller fordi det simpelthen
227 bare ikke kan lade sig gøre - men jeg synes måske... Det er den samme problemstilling jeg ender i, igen og igen...
228 Så snart at det ligesom er noget meget håndgribeligt, eksempelvis en lille designændringer eller det er "Kan du

229 finde ud af at skifte batterier i det her" så du'r alle brugere, fordi de kriterier der ligesom skal opfyldes, de kan
230 opfyldes af alle. Men så snart du ligesom kommer til det punkt at det skal være en slutbruger eller i hvert fald en
231 bruger med en hvis viden, så kan du ikke bare fodre dem en kontekts. Jeg havde eksempelvis det problem at, hvis
232 man lod som om det var diabetikere man ikke kunne få fat i, så bare det at fortælle dem at de ville skulle stikke
233 sig 3-5 gange om dagen, det virker af meget for en almindelig person, men for en diabetiker er det fuldstændig
234 acceptabelt, for det er de vant til. Så hvordan man ligesom får fodret noget kontekts ind, hvor det mangler uden
235 at man trækker data skæv og sådan noget. Jeg ved ikke om du har et bud på det eller om du mener det ikke kan
236 lade sig gøre?

237 **P:** Uden sådan lige og have tænkt særlig meget over det, så hælder jeg til at situationer som den der, dem kan du
238 ikke simulere. Du kan ikke tage nogle naive eller andre brugere som testdeltagere. Det trort jeg simpelthen ikke.
239 Altså ligesom... Jeg tror heller ikke, hvis du tester høreapparater, tage hørerne også at de skal forestille sig at de
240 er døve eller svagthørende. Jeg tror heller ikke helt man kan forestille sig at være stomi-patient uden øhm....

241 **F:** ... Ja at have stomi?

242 **P:** Ja.

243 **F:** Nu er det meget ekstremer, men simpelthen det her med at der er altså en grænse for, hvor meget kontekts
244 man kan fodre, uden at det vil skævvride resultater?

245 **P:** Ja, altså i forhold til at du nu siger kontekts... jeg tænker sådan...

246 **F:** Ja, eller erfaring-agtigt eller viden...

247 **P:** Ja, altså viden, følelser og holdninger er jo svære at putte ind i folk. Men du kan godt... hvad skal man sige...
248 Hvis det er en Epinephrine pen, kan du godt, i hvert fald for myndighederne, få lov at simulere stress på en anden
249 måde end at...

250 **F:** Ja, end at finde folk med rigtigt stress?

251 **P:** Det er mere det der med... Du ved, hvis du har set Pulp Fiction eller et eller andet andet, det er liv eller død,
252 hvis du ikke injicerer den her sprøjte ind i dig selv, jamen så dør du. Og hvis du skal teste det på børn, så er der
253 også noget etisk i forhold til det og hvordan siger du det til et barn og forstår de overhovedet det? så der er andre
254 måder at simulere et tilsvarende stress-niveau på. Det kan godt ske konteksten ikke er den samme, men der er
255 andre muligheder for at simulere et tilsvarende eller passende stressniveau på.

256 **F:** Ja okay. Så, hvad skal man sige... alternative tilgange kan overvejes i visse situationer?

257 **P:** Ja, jeg vil sige at de bør.

258 **F:** Ja, selvfølgelig, hvis nu eksempelvis med børn - det er selvfølgelig svært at gribe an og der er alt det etiske
259 som du nævner. Men det er måske svært og tilføre almindelig viden, eksempelvis det her med hvor mange gange
260 man skal stikke sig om dagen, uden at det virker overrumplende?

261 **P:** Ja. Ja det synes jeg.,

262 **F:** Jeg arbejder som sagt henimod det her med, om man kan substituerer den her population, men som du nævner,
263 i de fleste tilfælde, så er det bedre at smide de ekstra penge efter en bedre rekruttering. Tænker du det for kun de
264 summative altså de endelige tests, eller?

265 **P:** Nej, det tænker jeg netop ikke. Ikke for allei formative tests, men for mange fomative tests også. En ting er at
266 du kan... Hvis vi tager det her eksempel med fontstørrelser og at du så får Hr. og Fru Jensen ind til at bedømme
267 noget med fontstørrelser, så får du kun evaluert eller vurderet på det. Hvis du har rigtige brugere inde, jamen så
268 får du også noget rigere data og noget yderligere input, som du egentlig ikke har spurt om, men som kan vise
269 sig at være virkelig værdifuldt til at informere din designprocess. Så det er også vigtigt ikke bare i de summative,
270 for der kan du ikke påvirke dit design særlig meget, hvis overhovedet, men at få det ind i de formative faser, hvor
271 du rent faktisk har et design der er modtageligt for input, i forhold til og ændrer lidt.

272 **F:** Ja, så hvad skal man sige... At den egentlig målgruppe rent faktisk giver mere end man først regner med?

273 **P:** Ja, den giver mange gange mere end bare det du måler på og spørger om.

274 **F:** Ja og det er det jeg tænker. Jeg kommer jo ikke til at afleverer et resultat der siger du bare kan bruge Hr. og Fru

- 275 Jensen også regne med at du får det samme, men mere det her med at du er nød til at vide, hvilke begrænsninger
276 der ligger i og gøre det på den her måde.

Chapter E

Analysis of Interview

In this appendix two tables are presented. The tables consists of the Meaning Condensation Analysis for each of the two interviews conducted. The results from the analysis of the interview with Tina Øvad is presented first in Table E.1. Results from the analysis of the interview with Morten Purup is presented in table E.2. Each analysis was conducted by following Kvales Meaning Condensation Analysis, described in appendix B. Additionally, each analysis is based on the transcript, provided in the appendix D.

Meaning Condensation of Interview w. Tina Øvad

Table E.1: Meaning condensation from the interview w. Tina Øvad. The *Unit* is used for reference, hence **U1** in the report would refer to the first unit of the analysis. The *Theme* describes the essence of the natural unit and the *Condensed Description* is a concise description of the information found in the *Natural Unit*. The *Natural Unit* is the actual statement from the interview.

Unit	Theme	Condensed description	Line	Natural unit
1	Background for Scope of Thesis		03-12	F: Ja, altså jeg ved ikke, vil du have lidt baggrund på det, eller... P: Meget gerne! F: Altså, man kan sige jeg har brugt et halvt år ude i Novo Nordisk, i praktik, hvor jeg ligesom har lavet usability og user research. Jeg så ligesom en mulighed i enten at spare nogle ressourcer, i forbindelse med det her rekruttering og sampling af brugere, men jeg så også den udfordring, at når det er medicin som måske er til sjældnerte sygdomme, som også udvikles derude, så kan det være utrolig svært at få fat i den der population. P: ja... F: så hele mit speciale, går egentlig lidt ud på hvordan man ligesom tilgår sampling, når ens population ikke er i rækkevidde eller hvis ens produkt er så konfidenstilt, at det ikke kan komme ud af virksomhedsrammerne. Min vejleder nævnte så at du arbejder med noget, der var lidt lige sådan.
2	Situational sample criteria	The need for end-users isn't consistent through all forms of evaluation	13-19	P: Ja, altså det er.. Det er ikke helt så hårde omstændigheder, som du arbejder i, må jeg erkende. Men lidt på siden, er jeg i hvert fald i gang med at kigge på det her med, hvornår tester vi med hvilke typer folk, ikke? Fordi vi har jo, i forhold til vores arbejde at øhh - altså nogle gange bliver vi nød til at teste med potentielle end-customers, men nogen gange er det jo også human factors vi tester, nogen gange er det sådan noget lidt mere overordnet, hvor vi ligesom kan sige jamen altså det er ikke super relevant det lige er en ende-bruger der er med her og så er der især, hvis vi taler selve UX-delen af det, i forhold til os, ikke? Der bliver vi som regelt nød til at skulle have nogle fra vores kunde-segment ind og teste med det.
3	Analysis of population	In some cases perceptual and cognitive abilities might not influence results	20-29	F: Ja, det er lidt der... Jeg har jo så læst helt basiske teorier om sampling også ligesom også prøvet at finde noget empirisk gennem studier. Jeg synes ligesom der er den her opdeling af at enten bruger man sine end-users eller også bruger man nogle eksperter - og der er ikke rigtigt noget litteratur på, hvad man sådan kan gøre, hvis populationen er uopnåelig. P: Ja... ja fordi det kommer også lidt an på ikke, fordi jeg sidder lige og tænker i forhold til du siger det her med at det er medicin til sjældne sygdomme, ikke? F: Ja, det kunne være en case, ja. P: Ja, for eksempel ikke - eller det er netop det her med en population uden for rækkevidde, men jeg tænker jo stadigvæk, altså hvor meget er det er noget der er hængt op på den specifikke sygdom, altså hvor det er et eller andet der gør at de skal gøre tingene på en anden måde end man normalt ville gøre.
4	Hypotheses	Consider analyzing the need for each evaluation, to widen the criteria for sampling	30-37	F: Ja, ligenoagtigt. Det er den... Den løsning jeg har en hypotese om der kunne fungere, det vil være det her med at man går ind og analyserer hvad der egentlig er brug for i den enkelte test. P: Ja lige præcis. F: Så det er lidt hypotesen, men jeg har ikke rigtigt noget som helst der støtter op om det, andet end sund fornuft-agtigt.
5	Practical view	From a practical view, some types of evaluation diminishes the requirements of a representative sample	38-41	P: Ja, jeg må jo så også erkende at jeg slet ikke har kigget ind i det, fra et teoretisk perspektiv. Altså jeg har kigget ind i det fra et meget meget praktisk synspunkt der hedder - jamen altså vi kan jo gå ind og vurdere; er det human factors, så kan vi teste med mere eller mindre hvem som helst; er det UX-delen af det så tester vi med potentielle kunder, ikke?

Continued on next page

Continued from previous page

6	Potentiel outcome from thesis	Differentiation in sample criteria in relation to methods could form a matrix	41-45	P: også er der et eller andet range in between, hvor man så bliver nød til at graduere lidt, ikke? Fordi jeg tænker sådan lidt, altså umiddelbart i forhold til dit - altså jeg jeg bliver ved med, når du taler om det, jeg tænker sådan en eller anden matrix op, ikke? Hvor det er sådan noget med at i forhold til human factors, så er det den her type ting vi kan teste og det er egentlig lidt ligemeget hvem vi har inde, det er bare nogen hvor der ikke er nogen dissabilities også videre også videre.
7	Actual impact of criteria	Disassembling and analyzing the impact of sample criteria (persona-like approach)	45-51	P: Hvis at det er at det er noget, hvor der er nogle dissabilities, vil man så stadigvæk kunne finde sit sample i en anden gruppe end dem der lige præcis har den specifikke sygdom? I min optik, der tror jeg... Altså jeg hører det meget som at det kan skiller rigtig meget ad. Altså det her med, at det kan godt være det er en sjælden sygdom, men hvad impact har det på patienten? Hvis det ikke er noget der har impact på, netop som du også selv siger, med noget kognitivt eller noget fysisk af en eller anden art, ikke? Så umiddelbart tænker jeg det kunne være interessant at kigge på - hvordan skal jeg formulere det - en form, altså det er jo ikke en persona, men det er jo lidt derhenaf, ikke?
8	Criteria-based population substitute	Subtitue population to be defined based on criteria from targeted population	51-57	P: Ligesom så at sige, du har den her sygdom, det kan være X, [ikke? F: [Ja... P: og i forhold til X, der er der nogle kriterier det kan være at personen er lam i hænderne for eksempel, ikke? Eller det kan være... Jeg ved ikke hvad det kunne være, men det kan være mange forskellige ting også simpelthen stille de der kriterier op, der vil være ved de forskellige sygdomme, som kan have indvirkning på en interaktion. Også simpelthen caste... rekruttere efter det, ikke?
9	Design of Evaluation	Designing the evaluation to avoid or minimize impact of influencing factors	73-82	F: Ja og det fører fuldstænding ind i det jeg har indtil videre og man kan så sige det næste problem det er så så snart du går over, hvor det faktisk er de her lidt mere kognitive faktorer der gør sig gældende - jeg har rigtigt svært ved at se en mulighed i hvordan man kan substituere den her population, hvis man ligesom er ude efter de her lidt mere User Experience relaterede ting, hvor det ikke bare er håndtering og usability. P: Ja, men spørgsmålet er om man så skulle tage og vende den på hovedet også sige "okay" der er nogle kognitive ting der spiller ind, så de her personer vil ikke kunne x, y og z også vende den den anden vej rundt også sige vi ved de ikke kan x, y og z, hvordan er det så vi sørger for at de ikke har behov for at gøre x, y og z. F: Ja, så måske [...] hvad skal man sige. P: [Så i stedet for egentlig og hjælpe med at gøre x,y og z så skal man egentlig gøre sådan at hjernen ikke behøver at gøre x,y og z eller i hvert fald minimere x,y og z.
10	Learnings and experience	Know the limits of using a substitute population for sampling, in order to minimize effects of experience.	83-93	F: Ja, jamen helt sikkert, men jeg tænker også den retning der ligesom hedder i forhold til at når du bedømmer user experience så, i hvert fald inden for diabetikerenes verden, der har de jo typisk været knyttet til et produkt dagligt, henover en årrække. Når de så får noget nyt, der har de jo rigtig meget med i bagagen, i forhold til hvordan de bedømmer det nye. P: Ja. F: Og sådan som jeg ser det så er den eneste måde og gå ind og substituere noget som helst der, det vil være at man tager sig nogle forbehold for, hvilke begrænsninger der er, ved ikke at bruge de rigtige brugere. P: Ja... dut dut dut, jeg sidder sådan lidt og tænker, fordi det er netop det der med at hvis der er rigtigt meget læring inde, netop som ved diabetikere, og i forhold til de her insulin penne og de er vant til at stikke sig selv og det er lige ud af landevejen - men man kan også vende den om og sige, hvis man nu laver noget til den her målgruppe, hvor det simpelthen bare er ekstremt nemt og let at have med at gøre?

Continued on next page

Continued from previous page

11	No need for evaluating things already known	People experience and learnings can be taken for granted and used positively if the product itself can be categorized as 'easy to use'	95-101	P: Fordi man kan jo sige... Man kan jo gå langt ud, og man kan gå den der vej, hvor man gør det så simpelt at hvis du har noget erfaring med det er det piece-of-cake, men er egentlig reelt set også piece-of-cake, selvom du ikke har det. Det er jo så sværere for jer som virksomhed at gøre det, ikke? F: Ja, og [... P: [Vi ved jo at de har en eller anden form læring allerede, så der er allerede noget vi kan tage for givet, at det behøver vi ikke at have klaret altså med udmåling af hvor meget de skal have eller hvor ofte de skal tage det eller hvad det nu kunne være, ikke?
12	More iterations to achieve an evaluation on pair with the actual population	Splitting an evaluation into several iterations with different samples with different criteria can possibly be seen as a complete test, but with different people in each iteration	102-131	F: Ja, men hvad tænker du så, hvis det er et nyt koncept, og der ikke rigtigt er noget data på om det er simpelt eller ej? Fordi det kræver meget af designet at man bare kan lave et statement [om... P: [Helt klart, og det bliver nogle iterationer man skal ud i og virkelig prøve det af. Og det kommer virkelig an på det koncept man arbejder med, tænker jeg og hvor det er man er på vej hen. Hvis det er et produkt der... Man bliver nød til at starte et sted og nød til at ligge en baseline der siger hvor simpelt det skal så også være og hvor meget kan vi egentlig regne med der er af læring allerede i den setting. Som jeg ser det, så bliver det sådan noget mapping, hvor man virkelig får det mappet ud. Altså så siger man, vi har produkt A her, og vi ved at vores kerne-kunder de har den her erfaring også har de de her problemstillinger i forhold til, eksempelvis ledigt som du nævner, ikke? Altså simpelthen bare få mappet ud alt det der egentlig kunne have indvirkning også egentlig få lavet en fin sti, der så dækker så meget som muligt. Jeg tænker virkelig det er sådan noget med at man har en eller anden form for matrix, hvor man kan skravere noget af, også sige altså det her er dækket ved at teste sådan her, det her er dækket af at teste den her type personer, det her er dækket ved at teste dadadada også egentlig sige altting er testet, også summere op at det hele er testet, men så måske ikke med den specifikke, hvor det er den samme person der har kunne dække alle de krav, der nu en gang er. F: Ja, så lidt det... sådan som jeg forstår det, med sådan en matrice her, at du måske får lidt flere iterationer, fordi du er nød til at have nogle forskellige. P: Ja, det tænker jeg. Så simpelthen gå ind og sige nu er det alt det her med eksempelvis håndtering, så det kan være at man bliver nød til at teste med ledigt - så kan du sige i forhold til den fysiske interaktion, så er det egentlig dækket i forhold til den problemstilling. Så kan du sige der er nogle kognitive faktorer der spiller ind, jamen så enten lave det så nemt at det virkelig bare er lige ud af landevejen eller finde nogle der har nogle af de samme problemerstillinger, også kan du sige, så er det testet ved case 2. Så har du noget andet, hvor du så kan teste det ved case 3, også har du noget fjerde, hvor du så kan teste det på en eller anden måde også. Så nu er det alt sammen testet, men det er ikke med den samme person og samme målgruppe, men så må vi egentlig bare konkludere at fordi det faktisk virker i forhold til folk der har den her problemstilling, den her problemstilling og den her problemstilling så må vi kunne konkludere at produktet faktisk også vil kunne virke til dem det er designet til.
13	Splitting an evaluation into smaller fragments	Map out the criteria possible to test in a given context. Then do a number of evaluations that could represent a full study	133-137	P: Så jeg tror virkelig det sådan umiddelbart, vil jeg tænke, det vil være sådan et kort man tegner over det - få testet så meget som muligt, men så bare gøre det i forskellige steps, ikke? F: Jo, så studiet... Hvis man havde et studie til de rigtige brugere og sammenlignede med det man ville lave, så ville det man ville lave, det ville være mindre studier? P: Ja, også måske 5 af dem, men med forskellige 'foci' i hver.

Continued on next page

Continued from previous page

14	Awareness on context and evaluation objectives	Awareness of what information a sample is given is critical. It is however very based on the type and goal of the evaluation	138-159	F: I forhold til det her med, hvis de har erfaringer og læringer - sådan noget som kognitive walkthroughs og sådan noget, hvor man ligesom giver noget kontekts og selv sætter op, hvordan det burde være - tænker du at yderligere information til en bruger, som måske ikke kommer fra den rigtige population, at det ligesom vil være med til at give noget? P: Ehmm... Jamen jeg tænker næsten det modsatte, altså hvis jeg forstår de spørgsmål rigtigt? Fordi jeg tænker netop at der har i behov for at pumpe dem for så meget information som muligt? F: Ja, jeg tænker mene, eksempelvis, hvis jeg hyrer en helt naiv, som aldrig har haft med diabetes at gøres. Hvis jeg så fortæller ham en diabetiker vil typisk gøre det her 3-5 gange om dagen, tror du det vil give noget input, som ligesom er bundet lidt sammen med en erfaring, som en rigtig diabetiker vil have? P: Den synes jeg er svær, for et eller andet sted er jeg delt. Et eller andet sted tænker jeg det vil være en fin kontekst at sætte omkring tingene. Jeg ville være lidt bekymret for at jeg ville komme til at farve dem for meget, fordi jeg tror der vil være forskel på - for eksempel med diabetiker-eksemplet, der er du så vant til at du stikker dig fx. 5 gange om dagen altså hvor du ikke tænker over det, men hvis du siger det til mig, så ville jeg tænke "Hold da op, det er godt nok mange gange" og hold da op, så ville jeg have mange forbehold. Så jeg tror det er meget vigtigt det der med, at gå ind i hver også lige tage det der reality-check, hvordan ville det påvirke en at høre det, i forhold til en der er vant til det. Jeg tænker så også til gengæld, at nogle gange tror jeg måske godt det kunne give mening og gøre det. Men jeg tror virkelig det er en vurderingssag fra case til case, fordi jeg kan se nogle fordele, men jeg kan altså også se nogle ulemper. F: Ja, jamen helt sikkert. Jeg ser også ulemperne, men det er lige så meget det her med at fokus er at substituerer den her population, som man ikke kan nå, ligemeget hvad. Og det er sådan lidt, i mit hoved, det er det her med at man måske godt gøre nogle ting, eksempelvis give noget konteksts [... P: [Ja, så sætte scenen for det?
15	Additional requirements for analysis	Analyzing data from a substitute population requires more effort before, while and after the conduction.	160-172	F: Jo, men så samtidig så skal man også have i baghovedet, hvilke begrænsninger det ligesom giver. P: Ja. Og ja, jeg tror også du kan komme rigtig langt, ved netop når du har resultaterne, så sætte rammen omkring det og sige det er i den her konteksts det er blevet tal om og det er den her type information de har fået. Så begynder man også at kunne tegne sig lidt et billede af hvor man er på vej hen med det. Så jeg tror det er meget vigtigt det der med at kunne sætte den rigtige konteksts, eksempelvis hvis vi nu siger det her med stik-eksemplet igen, det er med folk der ikke er vant til at stikke sig og de har fået den her information og det er det her resultat der er kommet ud af det. Der er meget i analyse-arbejdet man skal tage højde for når man så analyserer det, ikke? F: Ja, så man kan måske argumentere for, at selvom det ikke kommer fra den rigtige population, så hvis der bliver taget hånd om det, både før og efter [... P: [Ja og stadigvæk have i mente, det her med at man også vil kunne farve dem i en retning af de ting man siger, hvis det er noget de ikke er vant til, hvor den population du egentlig gerne ville have testet på det er noget der egentlig er en del af deres dagligdag og deres daglige vaner.
16	Feeding of context	The influence of feeding additional context can implement biases in the evaluation	173-177	F: Ja, det er selvfølgelig rigtigt, det er bare også utroligt komplekst P: Ja, det er ekstremt komplekst! Jeg tænker også stik-eksemplet er et virkelig godt eksempel, for jeg tror en diabetiker vil tænke "Nåh ja", de er vant til det og det er en del af deres rutine, så det vil jeg slet ikke tænke over. Hvorimod, hvis du ikke er vant til det eller endda har nálefobi, eller hvad det nu er, så vil man nok tænke "shit mand - stik sig selv 5 gange".

Continued on next page

Continued from previous page

17	Deatiled mapping of sampling criteria for each type of evaluation	The possibility to use different methods for evaluations and samples could be valid, but the criteria of sample has to be mapped detailedly	178-187	F: Ja, og det er jo lidt de udfordringer, man sidder med tilbage. Når man sådan kigger tilbage på teorien og de studier der er lavet, så synes jeg bare det hele siger lidt, det er vigtigt at have den rigtige population, men samtidig siger det også lidt, hvis du begynder at sammenligne studie-design, eksempelvis field-evaluations og laboratorie-studier, at man kan godt afdække noget af det samme, men man skal igen bare være klar over hvad betydning konteksten har, for det man laver. P: Lige præcis, ja - og det er det virkelig. Det er også derfor jeg tænker at det mest brugbare vil netop være det her med... Virkelig lave det her... Altså et eller andet skema over... Okay den her target-group, hvad kendetegner den? Også ligesom begynde at dele ind; de har den her problemstilling - hvordan kan vi teste det og hvad er en alternativ rekruttering til det? I forhold til det her med alternativ - det er altså virkelig at få det delt op så man maksimere, hvor mange man potentielt vil kunne bruge i forhold til det specifikke studie.
18	Identity roles and personas	Identity roles and personas for "the context" e.g. type of disease.	188-195	F: Ja, altså sådan nogle identitetsroller? P: Ja, det er ligefør jeg tænker personer, fordi i stedet for at have en person, så er det egentlig en sygdom. Så kan man sige, hvis det er diabetes, hvad er diabetes; de skal i hvert fald have noget insulin. Så er det nogen gange skal vi teste med nogen der kan stikke - altså nogen der er vant til at stikke sig selv, ikke? Også hvad det nu ellers kunne være med diabetes, ikke? Simpelthen sige hvad kendetegner det og hvor vil vi kunne få den samme type person? Nej, ikke samme type person, men den samme øhhh.... Hvad kalder man det? F: De samme karakteristika der beskriver... [P: Lige præcis, ja lige præcis.
19	Usage of targeted population	Validation can't be done without the right population. Additionally the targeted population should be used early, to get a feeling of them	196-205	F: Ja, og så er det jo ligesom at arbejde sig frem til, hvad de her karakteristika er, og det er jo så både de perceptuelle og de kognitive. P: Ja lige præcis, og så finde alternativer til hver - også vil du højest sandsynligt støde på at der er noget du ikke vil kunne teste. Altså hvor du så netop så skal ind og sige "Nu mener vi at vi har gjort det så simpelt, at det burde kunne lade sig gøre", men hvor man reelt set ikke vil kunne validere det, medmindre du så får den specifikke gruppe ind. Der vil jeg så tænke at alle iterationerne, så prøv og se om du kan gøre det med nogen der... Det kommer lidt an på, for det kan også være man skal starte op med faktisk at have nogen inde, der har den specifikke sygdom, så man virkelig får noget føeling med, hvor de er henne, ikke? Men i hvert fald være meget påpasselig med hvornår mener du at det er validt at kunne teste med de andre og hvornår mener du så at det virkelig er vigtigt at du tester med den specifikke målgruppe.
20	Targeted population is necessary in some cases	Not all evaluations can use a substitute sample, but it is better to test them once, instead of everytime	206-210	F: Man kan sige, at nu har jeg selvfølgelig gjort det lidt svært for mig selv, ved bare at sige at den egentlig population er utilgængelig. Man kan selvfølgelig sige, at hvis man bare kan udskyde, sådan at man måske skal teste på dem én gang i stedet for 10 gange, det er jo også... P: Det er jo også... Ja det er bedre - meget bedre end ingenting og meget bedre end at skulle gøre det hver gang. F: Ja, lige nøjagtig.
21	Targeted population is necessary in some cases	Situations where the targeted populationen is the only solution exists. Either admit it or inform that not doing it with the targeted population is a "wild card"	211-215	P: Ja, for jeg tror virkelig det der med at få mappet det ud og ligesom kunne sige - der er nogle steder, hvor der er et alternativ - og når du så får det mappet ud, så er der også hvor du simpelthen ikke kan finde noget alternativ. Så må du så sige at enten så kan vi komme over det ved og så gøre et eller andet, hvor vi, med de kompetencer vi har, mener at det er klaret eller at der bliver du simpelthen nød til at - lige der - bliver du simpelthen nød til at skulle teste for at kunne være sikker eller også må det være 'Wild Carded' der er lige i den del af det, ikke?

Continued on next page

Continued from previous page

22	Possibilities of sampling decided by context	A sampling tool won't be able to be generalized, as sampling options are decided by context	216-222	F: Jo, og hvad skal man sige... Det der Wild Card eller de alternativer hvor man ikke kan teste, det vil jo så igen være situationsbestemt af eksempelvis diabetikere, de har måske nogle ting som er uopnåelige, men så snart du ligesom kigger på en anden population til et andet produkt, så er det måske noget andet? P: Ja. F: Så det her med at lave et generelt værktøj i forhold til at udvælge samples, det synes også lidt uhåndgribeligt, for det kan jo skifte fra gang til gang. P: Jamen præcis. P: Lige præcis, og jeg tror også at jeg er stødt på sådan et tool der en gang. Noget webbaseret, hvor man netop siger det er de her omstændigheder jeg arbejder indenfor, hvilken metode skal jeg så vælge. Jeg synes jo netop det er selve rekrutteringen og det her med at afdække hvordan man tester de forskellige requirements, det synes jeg jo lyder sindsygt spændende. Det andet der, det er sådan lidt en gammel traver, ikke? Men det der med hvordan mapper man ud, i forhold til at skulle teste et helt koncept, hvor man bliver nød til at dele det op i mindre tests for til sidst at kunne sige vi har valideret det store koncept med 5 mindre koncepter - så mener vi bestemt også at vi kan validere hele projektet, ikke?
23	Specific tool for sample alternatives	Only a useful tool, if it isn't the focused on methods	258-264	F: Nu ved jeg ikke, men hvor meget tid har du? For jeg kunne også godt tænke mig at vende det her med at hvis man har et produkt der er færdigt - eller har sådan et nogenlunde færdigt udseende-mæssigt, så lyder det som om at man lægger et sort klæde over også er adgangen til det ikke ret høj. P: Nej, det er vi ved at gøre op med - rigtig meget op med. Altså jeg mener der skal meget øje på... Det kommer jo an på, nu spørger du jo mig, ikke? Jeg siger jeg mener der skal så mange øjne på som muligt, også bliver vi nød til at sortere i de udsagn der kommer, og ligesom have den professionalisme der siger, altså i forhold til dem der udtales noget, hvilken baggrund har de og hvorfor bliver der sagt som der gør. Der er så også dem der siger vi skal ikke vise det til for mange mennesker, for de er for farvede og lalalala. Jeg er helt klart af den overbevisning at jo flere inputs du kan få, jo bedre, men du skal så også være god til at sortere i de inputs du får. Og altså jeg er helt klart fortaler for, at hvis du har noget, har billeder på din telefon eller et eller andet, hvis du har mulighed for at vise det til nogen, måske in-house, der har en NDA osv. ikke, for at få noget input, men selvfølgelig med det forbehold, der hedder at man skal tænke på, hvem de er og hvor de sidder hen og hvilken baggrund der er. Men der er altid - jeg synes der kommer mere positivt ud af det end der kommer negativt ud af det, i sidste ende. Om ikke andet, det gør dig skarp på at forklare konceptet og det gør også dig skarp på så ligesom at sige... Jamen nogen gange bliver der sagt noget der kan trigge dig. Det kan godt være det ikke er det specifikke der bliver sagt, men det er mere en eller anden lille kommentar der kommer, hvor du kan tage det op og sige "Gud, det har jeg slet ikke tænkt på, men gud der er da et eller andet her jeg bliver nød til at grave lidt mere ned i og lige få kigget på". Så altså jeg er meget stor fortaler for, ja... At vise det til så mange som muligt.
24	High exposure of products and professionalism	Keeping products in the business framework isn't sufficient. The more input, the better. One just has to be professional in sorting the inputs, based on the background of the people giving them.	267-284	F: Nu ved jeg ikke, men hvor meget tid har du? For jeg kunne også godt tænke mig at vende det her med at hvis man har et produkt der er færdigt - eller har sådan et nogenlunde færdigt udseende-mæssigt, så lyder det som om at man lægger et sort klæde over også er adgangen til det ikke ret høj. P: Nej, det er vi ved at gøre op med - rigtig meget op med. Altså jeg mener der skal meget øje på... Det kommer jo an på, nu spørger du jo mig, ikke? Jeg siger jeg mener der skal så mange øjne på som muligt, også bliver vi nød til at sortere i de udsagn der kommer, og ligesom have den professionalisme der siger, altså i forhold til dem der udtales noget, hvilken baggrund har de og hvorfor bliver der sagt som der gør. Der er så også dem der siger vi skal ikke vise det til for mange mennesker, for de er for farvede og lalalala. Jeg er helt klart af den overbevisning at jo flere inputs du kan få, jo bedre, men du skal så også være god til at sortere i de inputs du får. Og altså jeg er helt klart fortaler for, at hvis du har noget, har billeder på din telefon eller et eller andet, hvis du har mulighed for at vise det til nogen, måske in-house, der har en NDA osv. ikke, for at få noget input, men selvfølgelig med det forbehold, der hedder at man skal tænke på, hvem de er og hvor de sidder hen og hvilken baggrund der er. Men der er altid - jeg synes der kommer mere positivt ud af det end der kommer negativt ud af det, i sidste ende. Om ikke andet, det gør dig skarp på at forklare konceptet og det gør også dig skarp på så ligesom at sige... Jamen nogen gange bliver der sagt noget der kan trigge dig. Det kan godt være det ikke er det specifikke der bliver sagt, men det er mere en eller anden lille kommentar der kommer, hvor du kan tage det op og sige "Gud, det har jeg slet ikke tænkt på, men gud der er da et eller andet her jeg bliver nød til at grave lidt mere ned i og lige få kigget på". Så altså jeg er meget stor fortaler for, ja... At vise det til så mange som muligt.

Continued on next page

Continued from previous page

25	Informed design and Informed decisions	Prototypes should be shown in early stages, thus making it possible to use the input to change design and features. This helps obtaining informed design by taking informed decisions.	285-307	F: Der kan man jo så sige, at det er lidt at arbejde sig væk fra det der med at det er så konfidentielt at man ikke kan vise det uden for virksomhedsrammerne. P: Ja, og det er virkelig noget vi er ved at gøre op med. Vi havde vores første sådan brugertest i London sidste uge, hvor det er første gang vi selv er ude og lave det, altså med interne folk og hvor vi gør det selv. Vi har været ude og lave nogle småting med få personer, men her var det med 20 personer, - nej 18 personer til to koncepter, i forhold til portofolio og så videre. Bare det her med at få vist de prototyper der nu en gang er. Jeg er så til gengæld ikke fortaler for det her med high-fidelity prototyper, det har vi sådan en tendens til stadigvæk at gøre rigtig meget i. Men jeg kan godt lide det her med, at man har sådan nogle idéer, nogle tegninger eller nogle renderinger eller et eller andet, så man så kan snakke ud fra det og ligesom bare høre hvad folk har at sige. Jeg har også den der meget med at jeg går meget ind for informeret design og informerede beslutninger, men det kan godt være vi ikke beslutter det folk siger, altså også når vi laver brugertests, det kan godt være at folk de siger de vil have - altså vi viser dem A, men så siger de, de vil have B, men så kan vi godt gå ind nogle gange og sige, det kan godt være de siger det, men vi vælger faktisk stadigvæk at gøre det andet. Fordi vi mener vi bestemt har en bedre case, men så er det en informeret beslutning vi træffer, så vi ved når vi så lancere at så kan der være et problem her, fordi folk har altså gjort opmærksom på det. Jeg kan i hvert fald mærke, når det er den approach vi tager - og det er meget meget nyt, så vi har ikke noget ude, hvor vi har gjort det sådan endnu. Så jeg kan sagtens sidde og spille smart omkring det, men jeg kan også høre det gør også designerne mere trygge, fordi de stadigvæk har et mandat, fordi jeg har lavet rigtig meget på at vi skal satse på at være så brugercentriske som muligt også står der nogle koncept-managers og nogle designere på den anden side også siger "Nej, nej, nej det skal være design-driven det hele" også siger vi jamen okay, men hvor kan vi så mødes? Jamen vi kan godt mødes halvt her, altså næsten på midten, hvor vi så siger, det er fint vi tester det, og det er fint at vi får nogle rigtig gode findings på alle tingene, men det er ikke sikkert at vi hører efter hvad der er blevet sagt. Men det er mere informeret, ikke?
26	Fidelity-based issues in design of evaluation	When designing an evaluation the fidelity of the prototype has to be taken into account, as some aspects can't be covered, depending on the fidelity	312-319	P: Nej præcis, og det begynder lige så stille at løsne op omkring det nu og det er meget mere det her med at vi lige går ned og kigger på prototyperne, lige snakker om det og har folk med dernede og sådan, ikke? Men det er noget der tager ekstremt lang tid, også det her netop med at gå og vise tingene og få noget feedback, for der er meget af det her med "Nej det du'r ikke", fordi folk kan ikke finde ud af at forholde sig til det, hvis det ikke er en påen prototype - og det er jo sådan lidt, det kommer jo an på hvad vi snakker om, fordi hvis det er interaktion, så gør det jo ligemeget, altså det der vi spørger hvor mange et eller andet vil I have på noget, for det kan folk godt forholde sig til. Noget andet er at forholde sig til om de vil have det specifikke produkt ind og stå i deres hjem, fordi hvordan det så helt ser ud, det er helt industriel design-agtigt.

Continued on next page

Continued from previous page

27	Sampling to include other aspects of evaluation	Sampling is connected with so many aspects, hence a suggestion for developing a tool taking several aspects into account e.g. fidelity, method, design-maturity, population	325-340	F: Nu nævnte du den her matrix igen, tænker du at man måske... den vil jo nok være 2D, men tænker du 3D vil være en mulighed i forhold til at man måske opstiller sådan en matrice, også alt efter hvilken side du kigger på den fra, så tager den ligesom [... P: [Det kunne den sagtens, for jeg tænker virkelig.. Hvis nu vi taler ud fra participants, så har vi noget med hvilke krav er der fra... Eller hvilke requirements er der fra kundesiden og hvordan tester vi. Så må der også være noget omkring hvilken fidelity der så også skal være. Jeg tænker også tid er et aspekt i forhold til hvor moden et koncept er, men jeg tænker mere eller mindre det ligger i fidelity. F: Ja, og det ligger nok også i valg af metode i forhold til testen, en gang i mellem også - så det er måske implicit. P: Ja, men spørgsmålet er jo faktisk så, hvis vi tager det der træ, kunne man lave den sådan en 4'er, hvor man så siger, vi har nogle requirements, vi har... en metode... metode er der ja, også har vi fidelity der... for så har vi, hvis vi nu har metode inden også ikke, fordi så kan vi jo se, hvis vi så vender den på... så siger vi at metode - vi kan lige som sige at vi har den her metode og vi ved at det er det her vi vil, vi ved at det eksempelvis er en fokusgruppe vi vil holde, fordi det har vi mulighed for lige nu. Hvilk type requirements er det så vi vil kunne svare på, i forhold til det og hvilken type fidelity er det så vi skal bruge til det i forhold til. Vi kan også have nogle requirements der siger vi skal teste med folk, der er vant til at stikke sig selv. Der er et eller andet her, der godt kan blive til noget, tror jeg. Den skal lige gennemtænkes.
28	A need for investigating the markets needs and opinions	There's a need for getting input from the market. In-house decisions are helpful, but can't remedy everything	348-360	P: Forhen der havde vi nogle koncept-lead og nogle designere, der troede de var guder og vidste hvad folk vil have. Et eller andet sted tror jeg vi har været heldige, at vi nogle gange har haft folk der har haft flair for det de lavede, og flair for kunder og flair for usability og UX. Så har vi haft nogle gange, hvor det ikke har været tilfældet. Så det har virkelig været sådan en rutsjebane. Men.. meget stærke personligheder, der har fået nogle fede ting trumfet i gennem, men det har ikke været med fokus på usability og UX. F: Nej, på brugerden? P: Nej, slet ikke og det er netop det vi er ved at gøre op med nu. Vi bliver nød til lige at høre efter hvad markedet - eller om end ikke andet have en idé om hvad markedet mener om de ting vi laver. Men det er jo mange forskellige ting. Der er mange forskellige initiativer, der er ongoing lige nu og der bliver rykket rundt hele tiden og der er ændringer og der er organisatoriske ændringer og der er ændringer for hvordan vi arbejder. Vi er i gennem en agil transformation samtidig med - altså der er rigtig meget der rykker og nogle gange så tænker man, hvis man kan holde boldene i luften, så bliver det virkelig fantastisk, men nogle gange tænker man også ej ej, det er alt for meget på en gang, det hele må falde til jorden på et eller andet tidspunkt.
29	People before Technology	People before technology instead of user's having to adapt to something decided by a designer	360-364	P: Der sker rigtig mange ting, synes jeg og netop især på området med brugerinddragelse og... vi snakker rigtig meget people before technology, fordi det har vi været slemt til lige som at sige at det var folk eller vores kunder skulle egentlig bare indrette sig efter den teknologi der nu en gang var og det design vi havde valgt for dem. Vi bliver nød til at vende den her tankegang om, også sige hvem er de folk vi udvikler til og hvad er det egentlig de har behov for.

Continued on next page

Continued from previous page

30	Differences in segments, depending on perspective	Segments might change depending on perspective i.e. marketing or UX perspectives. Archetypical segments might be a solution.	366-372	P: Øhm, et eller andet sted, ja lidt. Vi er i hvert fald i gang med meget at få mappet ud, hvilken kundegrupper det er vi primært designer til også sørge for ligesom at få målrettet vores produkter i forhold til at nu er det det her segment, hvad er det så der er vigtigt for det segment? Så tester man med det segment, men stadigvæk har også øjnene åbne for at det også er andre end det segment der skal købe vores produkter. så der er mange forskellige ting, også er det hele det der med segmenter, er der forskel på om man snakker marketingssegmenter kontra de segmenter vi udvikler til? Det burde der ikke være, men det er det tit alligevel. Vi vil jo gerne have noget der er mere persona-rettet hvor de jo har noget der er mere arketyptisk rettet. Så der er mange ting der er lidt det samme også alligevel ikke helt. P: Altså der har været noget - men jeg tror også det har været meget person-bestemt og meget ad hoc også har det også været meget med 'måske bliver der hørt efter, måske gør der ikke - vi ved det ikke helt'. Altså jeg kan sådan se, jeg har været her lidt over et år nu, og vi har nogle produkter, hvor at de ligesom har rejst et skrig, hvor jeg sådan har sagt, at der er et eller andet, som ikke er helst som det skal være her også finder jeg ud af at der faktisk har været lavet brugerundersøgelse omkring det, der er bare ikke blevet hørt efter de findings, der så kom ud af det. Så der har været lavet, men det er som om der ligesom er blevet sagt "Ah, hvad ved de? Ah, vi gør bare som vi plejer". P: Men nu er tiden også ved at være mere moden til det. Da jeg lavede min P.H.D og lavede nogle interviews med forskellige virksomheder, der er også den der med at det kan godt være at folk eller virksomheder siger, at vi skal have fokus på usability og UX, men det er faktisk først når man virkelig har haft et produkt der er fejlet bigtime, hvor man så får øjnene op for at det kunne vi faktisk godt have fanget, hvis vi havde gjort noget i den retning. Det tror jeg måske, ikke fordi vi har haft nogen produkter der har været sådan, men nu er tiden også ved at være moden til, at virksomheder og forretninger kan se at de bliver rent faktisk nød til at inkluderer en eller anden form for slutbruger i det arbejde der bliver lavet.
31	Ignorance of user input and subjective interpretations	Due to the lack of US and US evaluations some problems have been discovered after release, as a result of keeping the design in a business framework	382-388	
32	Maturity of UX and US	The time for UX and US is now. Previously UX and US was eye-opening if a product failed miserably and the problems could have been solved by input from users in the earlier phases	395-401	

Continued on next page

Continued from previous page

33	Recruitment perceived as a constraint	Recruitment is somehow perceived as a constraint by many people, and it is hard to identify exactly why the recruitment process is seen as a constraint.	403-419	P: Det er generelt. Jeg synes generelt usability user maturity er... Jeg havde faktisk et møde i går, også med en, hvor vi snakkede om det - der har været meget snak om det, men ikke så meget gjort omkring det, hvor jeg tror at nu bliver det stille og roligt også en del af nogle processor, fordi det er sådan vi vil gerne og alle kan se idéen i det, men så når man har travlt og forskellige ting, så er det lidt det der bliver udladt og så er det sådan lidt på gefühl og adhoc også gør vi det lidt hist og her. Jeg synes at der er lidt en tendens til nu at der begynder at komme lidt mere process op omkring; hvornår gør vi det så og hvordan gør vi? Hvor sådan et tool - hvis du fik lavet det her, det ville også været helt ideelt. Hvis man så kunne adoptere det til andre virksomheder også. Og hvis man bare kunne sige vi har den her kuge eller det her sheet - også kan man egentlig bare selv begynde at putte noget ind i den, ikke? Helt ideelt, også fordi så bliver det lige pludselig ikke så farligt. Jeg kan mærke her, at det er rigtig meget rekrutteringen, der virkelig er et stort problem ofte, ikke? Jeg arbejde ved Radiometer før jeg kom her, hvor det også er medical devices lidt lige som Novo - ikke helt som Novo, for det er mere medicinsk - men også hvor det var meget snævre target groups, hvor det var svært at finde testpersoner, man bookede gennem bureauer, det tog en krig og det kostede en bondegård osv.. Hvor jeg egentlig havde troet at det ved B&O ville være let, fordi vi har kunder. så var der en business-model der sagde at man ikke måtte snakke direkte med kunder og dit og dat, så den har vi så også fået gjort op med nu. Men stadig den der med at det tit er rekrutteringen der er hæmskoen og jeg kan ikke helt lure om det er fordi det er det der er lettest at sige eller om det faktisk reel er dét det er. Jeg ser den virkelig som en hæmsko.
34	Willingness to participate is not an issue	Users and in-house employees are often willing to help. The constraint seems to be more practical and a solution would be to improve the accessibility of users.	420-437	P: Ja, men det er så det der med at der er nogle specifikke kriterier og det er et specifikt segment du gerne vil teste med, ikke? Så har vi den problemstilling, men så kan man så sige, hvad så hvis vi får det afdækket, hvad er det så der gør det? Det er så her kæden den hopper lidt af, fordi der burde virkelig ikke være noget der ligger til grund for det. Også her ved B&O, vi kunne jo reelt set bare tage folk ind fra gaden - eller til noget af det ville vi kunne. F: Ja, lige nøjagtig. P: Det undrer mig bare, at det ikke er det der sker. Der er en barrierer af en eller anden art, hvor det skal gøres let tilgængeligt. Jeg har været i gang med, nu er det så på hold lige nu, men netop med at lave sådan en participant pool, hvor det simpelthen vil være let og så ligesom kunne sige at det er folk der er interesserede i at deltage og som har signet sig op også vil man kunne skrive en eller anden fællesmail ud; "Hey vi tester i Struer mandag, tirsdag, onsdag - her er kalenderen, i kan bare booke jer selv hvor det passer". Det kan jeg mærke at det gør det lidt lettere, vil jeg sige. Men generelt er rekrutteringen virkelig en hæmsko i det her arbejde og der er også et eller andet - jeg ved ikke om det er noget med at folk er lidt generte og det der med lige at få spurgt. Generelt er folk - også internt - villige til at sige "okay, jamen jeg kan da godt lige komme og en test her og nu".
35	Time-based constraint	Sometimes people won't have time for participating	437-439	P: Det er ikke alle steder det er sådan, der hvor jeg arbejde før, der var det sådan "Nej, det har jeg sku da ikke tid til, hvad fanden tror du?". Så der er et eller andet der, men jeg er ikke helt sikker på endnu, hvad det er - men jeg synes at jeg begynder at kunne se lidt en tendens deromkring og at det tit er der hæmskoen er.

Continued on next page

Continued from previous page

36	Concerns for others	A concern for wasting people's times (especially in-house) due to bustle.	441-444	P: Ja, og der er mange aspekter i det, men den helt lavpraktiske er at man er bange for at man tager folks tid, har folk tid, folk har også travlt og dit og dat - det er en del af det. Der er bare også noget andet i det end det - jeg kan ikke helt pinpointe det endnu, men jeg kan begynde at.. De snakke jeg har med folk, hvor jeg kan høre det - når nævner de ting, så de statements der kommer de supporterer ligesom det der - så der er et eller andet.
37	UX and US researchers are prima donnas.	Researchers within the field of UX and US might be prima donnas. We've been taught to do exactly as the method states, and the possibility for relaxing on some criteria often doesn't exist. This lead to a way of thinking stating "If criteria are not met, it's not worth doing."	446-456	P: Ja, lige præcis og jeg kan også høre at det jo er mange steder det er sådan og det er virkelig underligt hvorfor! Så snakkede jeg også med nogle i går netop med - jamen er det også fordi at vi vil gøre det så godt og vi er bange for at fejle hvis vi rekrutterer interne folk? Fordi jeg tror også der er meget den der med at vi UX'ere er lidt nogle primadonnaer en gang i mellem. Vi har lært at man gør det på den her måde, så det er ved gud at vi gør det på den her måde. Så hvis vi skal til at begynde at slække på noget... nogle gange tænker jeg også det er bedre at have én end ingen og teste med og det er også nogle gange bedre at ligesom at sige, så tester jeg den her lille del af det. Det kan godt være det ikke er det hele jeg tester, men det er det er den her lille del også ved jeg i hvert fald at det her - det virker. Så tror jeg også nogle gange at vi har en tendens til at sige, hvis vi ikke kan teste det hele, så kan det også bare være ligemeget, ikke?
38	Mismatch between academic and business'	The academic methods for conducting user experience research and usability studies isn't conforming with the needs within a business. The relaxation on criteria might not lead to results categorized as academically valid, but still valid enough to be the basis of an informed decision.	457-473	P: Jamen netop. Den hører jeg rigtig meget og jeg synes især også - når jeg har arbejdet meget med agilt UX, de akademiske metoder der er, i forhold til at lave usability og UX tests, de er meget meget tunge og de er meget akademiske, det er noget der tager lang tid og det er noget der kræver relativt mange ressourcer også. Så det der med at få metoderne lavet lidt mere agile og lidt mere lette, det er altså også vigtigt. Det var det jeg arbejdede på med min egen P.H.D. Hvordan kan man modifcere de her metoder - det kan godt være det ikke er validt fra et akademisk perspektiv, men det er i hvert fald nogle findings der er valide nok til at kunne træffe nogle beslutninger ud fra. F: Ja, lige nøjagtig. Så længe man bare har lidt information at træffe sine beslutninger ud fra og at man ikke bare gør det på gut-feel. P: Lige præcis. Det arbejdede jeg jo så med i 3 år og jeg underviste også i det, hvor jeg kunne høre at det er noget der resonere rigtig meget med de virksomheder, der sidder og arbejder med det. Det er den der med... Så tager man den der dybe indånding og tænker "så skal det også være det der stooore user study" i stedet for det bare... Det behøver det jo ikke, det kan jo bare være 5 minutter og en håndtegning, hvor vi lige gennemgår noget og hvorfor. Så kan man sige "Ahhh det der giver simpelthen ikke mening, fordi det er ikke sådan man normalt ville gøre det og det her er altså indutristandarden indenfor... det er jeg da vant til" - også ser det som en finding og at det ikke bare er ligegyldig information, for det er det jo ikke. Og også italesætter det som at det faktisk er en eller anden form for brugerstudie.

Continued on next page

Continued from previous page

39	A paradigm shift in UX and US research is on the way, but not complete	The perception of user studies is slowly changing. The change in paradigm is that the categorization of user research is a sliding scale and not a cemented method of doing research	474-484	F: Jo, at man ligesom i bredere grad acceptere at brugerstudier kan gradbøjes og det ikke er... P: Lige præcis! Der tror jeg också der er et paradigmeskifte - jeg tror det er i gang, men jeg tror ikke det er helt fuldendt endnu. Det er ligesom at man kommer fra 'Det er sådan her man gør, eller också gør man det ikke ordentligt också gør vi det ikke' också til 'Vi kan också bare gøre sådan her'. Lidt ligesom det her med netop den her kuge eller hvad det bliver til. Hvor man kan sige det kan godt være det ikke er det hele der er testet, men det kan være det er 75% der er testet också har vi så en marginen på 25% der ikke er testet, men hvor vi enten konkluderer at det burde ikke være noget problem eller vi ved der kunne ske at hænde fejl, når den ender ved den rigtige slutbruger. Men netop det der med at kunne differentiere i det; det er ikke 100% men så er det 75% eller hvad det nu kunne være. Hvis man kan teste 75% med folk der ikke er i target gruppen också de sidste 25%, hvis de skal testes, så er det der man bruger de folk der kan bruges til det. Et eller andet sted, så er det jo sund formuft, ikke?
40	Vouching for valid results obtained by non-academic methods	Even though UX and US can be seen from many perspectives, it can be hard to go up against "the norm" and jeopardize one's competences and professional background vouch for valid results, obtained through a non-academic method.	485-489	F: Jo, lige nøjagtig, men jeg tror også lige så meget at det er at få overbevist folk om at man godt kan bruge den her sunde fornuft og at det ikke altid skal være som det står skrevet op og ned. P: Lige præcis. Det er virkelig også... Jeg prædiker meget om det og det resonere også fint, men der er stadig lidt en barrierer omkring de ting der - er det så validt nok og tør jeg med den baggrund eller position at stå inde for at det er validt nok, når jeg ikke har gjort det 100% som jeg er blevet lært?
41	The essence of studying possible alternatives within sampling		498-500	F: Det tror jeg det er det der skal til, for at komme fremad, for lige nu er det bare, hvis du ikke har den rigtige brugergruppe, så kan du lige så godt lade være. P: Ja, og det er jo fuldstændig skørt ikke?

Meaning Condensation of Interview w. Morten Purup

Table E.2: Meaning condensation from the interview w. Morten Purup. The *Unit* is used for reference, hence **U1** in the report would refer to the first unit of the analysis. The *Theme* describes the essence of the natural unit and the *Condensed Description* is a concise description of the information found in the *Natural Unit*. The *Natural Unit* is the actual statement from the interview.

Unit	Theme	Condensed description	Line	Natural unit
1	Scope of Thesis		07-19	F: Så jeg er ved at skaffe lidt empiri på hvad der egentlig rør sig i andre virksomheder og snakke med folk der ligesom har arbejdet med det. Man kan sige at det problem jeg så til at starte med, er at der bliver kastet rigtig mange ressourcer efter alt det her rekruttering og at man skal nå de rigtige brugere. Og det næste er så det her med at når man designer til en population, som stort set ikke kan nås - som måske gør sig gældende ved sjældne sygdomme - hvordan får man så ligesom testet det man er i gang med at designe? og det er nok mest det device-baserede jeg kigger på. F: Det jeg så har arbejdet hen imod er at hvis man har en eller anden type evaluering - lad os sige field evaluation - og man ligesom har sin test klar, men at man ikke kan nå den her population, hvad der så kan gøres af tiltag for at få nogle resultater, som stadig er valide. Så til at starte vil jeg gerne høre hvilke problemer der gør sig gældende i forbindelse med det du arbejder med. Jeg ved du har været ved Radiometer tidligere...
2	Bad experiences in previous evaluations makes the budget for recruiting unquestionable	Previous use of non-representative users resulted in findings not valid enough in order to conditionally ensure a safe and effective product to the authorities. Therefore, there's no questions asked regarding the expenses related to recruitment.	20-30	P: Ja... Jamen det har jo også været ret snævre brugergrupper vi har skulle rekruttere nogle gange, hvor man også er rendt ind i problemer, som følge af ikke at have brugt de rigtige testdeltagere. Så derfor blev der, i hvert fald til sidst i den tid jeg var hos Radiometer, ikke stillet så meget spørgsmål til rekrutteringsudgifterne til at få de rigtige testdeltagere, altså nogle der har en hvis erfaring med en bestemt type apparatur eller ja... specifikke behov osv. Det er simpelthen fordi man har kunne se de tidligere problemer man har haft ved og rekruttere for brede brugerprofiler fremfor nogen... Simpelthen fordi der har været og er for store forskelle. Hvis du bare rekruttere sygeplejersker der er van til at håndtere blodprøver - der er kæmpe forskel på om det er hæmatologi eller om det er blodgas analyser for eksempel. Vi har bare fundet ud af at vi simpelthen bare ikke kunne bruge resultater fra for eksempel hæmatologi-sygeplejersker og de har ikke været repræsentative for de intenderede brugere som vi skulle ramme. Det har været et problem i forhold til egentlig at kunne konkludere - kunne claime - at vi havde et sikkert og effektivt apparat betingesmæssigt overfor myndighederne.

Continued on next page

Continued from previous page

3	Goal-based sampling for evaluations	Depending on the goal of the evaluation a different variety of samples can be used	31-42	F: Ja, hvad med... Nu siger du myndighederne, det lyder meget som en summativ evaluering - hvad med de mere formative studier, hvordan har det set ud der, i forhold til sampling? P: Jamen det har været alt fra at vi tester på to ved nabo-skrivebordet eller nogen man finder i en anden afdeling, finans eller HR afdelingen som intet ved om, hvad der foregår i udviklingsafdelingen. Så det kan være dem, det kan også være... På radiometer havde vi laboranter internt, så det er jo egentlig super-brugere som kan repræsentere. Nogle gange ved de jo for meget, men... Ja det kommer jo an på hvad man tester, ikke? Fordi hvis du tester noget med om det er en stor nok skriftstørrelse på vores GUI, jamen så kan du tage Fru Jensen fordi de perceptuelle egenskaber og de forudsætninger hun har for at læse, er ikke anderledes end andre sygeplejersker, da det er basalt ens for de brugergrupper. Men hvis det er noget, der typisk har med det kognitive at gøre, noget man skal vide i forvejen eller noget man skal være særligt obs på eller noget der har med specifikke regler eller problemløsning eller omsættende information at gøre, så gælder det simpelthen om at få nogen der ved hvad de har med at gøre. P: Jeg vil mere sige det kommer an på hvad du undersøger, fordi det der med hvor langt du er i processen det er... Der kan være behov for at ramme lige præcis den intenderede brugergruppe tidligt i processen, men også... Ja, så tidligt i processen, at du stadig kan lave noget om. Jeg tænker eksempelvis i forhold til user needs og identification, der er det vigtigt at have de rigtige. Så kan der være noget i de aller tidligste faser, hvor man lægger sig fast på noget koncept, hvor det også kan være problematisk, hvis ikke man har den rigtige brugertype.
4	End-users in specific phases of development	The need for actual end-users can appear in most phases of the development, as it is the goal of the evaluation that dictates the kind of sample	49-53	P: Jeg vil mere sige det kommer an på hvad du undersøger, fordi det der med hvor langt du er i processen det er... Der kan være behov for at ramme lige præcis den intenderede brugergruppe tidligt i processen, men også... Ja, så tidligt i processen, at du stadig kan lave noget om. Jeg tænker eksempelvis i forhold til user needs og identification, der er det vigtigt at have de rigtige. Så kan der være noget i de aller tidligste faser, hvor man lægger sig fast på noget koncept, hvor det også kan være problematisk, hvis ikke man har den rigtige brugertype.
5	Substitute samples can be valid	Substitute samples can be used in different occasions	60-64	P: Ja... Jamen selvfølgelig er der områder, hvor du kan bruge andre personer og det... Så er der det her eksempel med legibility eller om fontstørrelsen er stor nok eller lysforhold og sådan noget. Der kan man bruge andre. I forhold til, måske, præference af noget look-and-feel kan det også være man kan benytte nogle andre.. i forhold til navigation i en software-GUI kunne det også være man kunne benytte nogle andre, og sådan noget som... Altså sådan noget som () og alt det antropometriske.
6	Experience and knowledge of users	Know-how knowledge was identified as a pitfall for using non-representative users	65-71	F: Ja - I forhold til de problemstillinger identificeret ved at bruge for brede samples, hvad var de typiske problemstillinger så? P: Jamen dels var det at de simpelthen ikke kendte til hvilke faldgruber der er ved en blodgasanalyse. Der er det man - det FDA kalder knowledge tasks - der var simpelthen nogle steps inden man analyserer blodprøven, som hvis du ikke udfører dem eller udfører dem i en forkert sekvens så udgør de en risiko. Der var også udfordringer i forhold til at protokollen, testprotokollen, havde statet specifikke inklusionskriterier til testdeltagerne, som ikke blev efterlevet og det var sådan noget som kvalitetssikring og (registrerings...) opponerede meget imod.
7	User training	Training users is not realistic in all scenarios	72-76	F: Ja. De mangler jo så lidt den her læring og konteksts omkring brugen af jeres apparat. Tænker du det er noget man måske kunne have frontloadet i forhold til og gennemfører en test og få valide resultater derigennem? P: Ja, altså alternativet er at du skulle oplære naive testdeltagere til at have samme viden, som repræsentative brugere - det tror jeg bare ikke er realistisk, ikke i det her tilfælde i hvert fald.

Continued on next page

Continued from previous page

8	Analytically approach	When possible an analytically approach can be taken. The goal is to break down task into discrete steps and analyzing them. The analytical approach should be paired with normal usability tests and other empirical data collections, as the two method compensates and replaces eachother	86-105	P: Øh ja. Den metode vi har fået mest ud af er egentlig task-analysen. I bund og grund kan du sige, der er ikke nogen grund til at gå ud og teste ting, som en mand kan finde ud af bag skrivebordet. Så vi benytter de analytiske metoder, som brainstorm og task-analyse som... altså vi benytter dem på den ene side også de empiriske, som usability test og det at indsamle data fra rigtige brugere til at kompensere hinanden. Til at nå begge veje rundt, ikke? F: Ja. Hvad siger du, task-analysen er det simpelthen hvor du går ind og kigger på den enkelte opgave eller hvordan forholder det sig? P: Ja, altså du tager hver task og bryder ned i diskrete steps, for eksempel hvis det er et eller andet apparat så kunne det være noget med at tænde det, eller at klargøre det også kunne det være noget med måle også kunnen det være noget med at få resultatet videre eller ud eller sendt til et IT-system. Det kunne også være for en forbrugsvare, du skal finde den her forbrugsvare, så skal du aktivere den også skal du installere den, Lidt ligesom en printerpatron eller sådan et eller andet. F: Ja og der får i så bare en... hvad skal man sige, en ekspert til at??? P: Ja, altså det kan laves af en person også kan det selvfølgelig reviewes af andre. Så er det jo for så vidt muligt også at få tænkt konteksten ind i det, lysforhold, lydforhold, afbrydelser... Hvis det er i en ambulance, så kan det være de kun har én hånd til at interagere med devicet med, fordi den anden hånd brugere de til at holde fast med, mens der køres stærkt. F: Ja, så det vil være lidt det der, hvis du ikke tog højde for det, i dit laboratorie-setup, så ville du ikke få valide resultater, eksempelvis? P: Ja for eksempel. Task-analysen er så den analytiske metode, så man kan identificerer noget af det her, inden man går ud og laver usability-tests.
---	-----------------------	---	--------	---

Continued on next page

Continued from previous page

9	Unpredictable situations can't be found through analytical processes	Task analysis can't sufficiently explain everything that could happen in a usability study, as unpredictable observations are often made, when real users interact.	107-128	F: Ja - i forhold til det her med kontekts, så synes jeg - nu har jeg kigget en del studier i gennem, og der er blandt andet nogen der sammenligner field evaluations og laboratory settings og det her og jeg synes det de kommer frem til er, at ca. halvdelen plus det løse kan identificeres i gennem de forskellige ting, men så er der en spredning på, hvad der bliver identificeret gennem de forskellige metoder. Eksempelvis field er der meget kontekts og laboratory, der er det nogle andre forhold der gør sig gældende - måske det at de sidder i en stol og gennemgår det. Tænker du at der er nogen work-around i forhold til det - altså... Mange af de samme usability problems bliver egentlig identificeret gennem de forskellige metoder for eksempel field, lab og heuristics og efter ca. halvdelen, så begynder det at være lidt mere spredt ud - så er det nogle forskellige problemer de får øje på. Tænker du at der er en work-around der kan gøre op for det eller tænker du at metoderne ligesom skal komplimentere hinanden i forhold til det? P: Jeg tænker at de komplimentere hinanden, for der er nogle ting i en usability test, som du bare ikke kan tænke dig frem til. Vi havde for eksempel... jeg kan lige komme med et eksempel først, også kan jeg komme med hvordan man måske kan komme omkring det analytisk. Vi havde et eksempel hvor.. for at måle på en blodprøve, så skal man trykke sprojen hen på et inlet - altså væk fra sig selv, man skal simpelthen trykke inlettet op. Der sidder sådan en lille fjeder der aktiviserer opsuget. Den her action eller den her sekvens, den er også illustreret med en video, hvor man ser en videosekvens af hånden der gør det her. Til at starte med så den her video-sekvens den loopede, så vi så simpelthen i brugertesten at folk de stod og trak hånden frem og tilbage, fordi videonj loopede. Så stoppede vi så videon, efter den var færdig og gjorde den grå og satte sådan et replay-ikon på. Det replay-ikon, det var på det tidspunkt hvor YouTube havde sådan en cirkel - sådan en pil rundt i en cirkel. Vi så så en testdeltager stå foroverbøjet over det her apparat og forsøgte at skru sprojen på - simpelthen som om der var et gevind. Vi spurte så, hvad tænker du nu, eller hvad prøver du på? siger så "Den siger jeg skal skru den på". Det havde vi jo slet ikke forestillet os gennem brainstorm og task-analysen.
10	Mental models as an expansion of task analysis	Investigating mental models could help overcome the unpredictable behaviors in an analytical way.	128-134	P: Der hvor man måske ville kunne identificerer det i de analytiske metoder, det vil være ved at kigge på hvilke mentale modeller har folk og hvilke mentale modeller udstråler designet, eller ikoner osv. F: Så hvad skal man sige, lave ligesom et bredt sprektre af hvilke mentale modeller der kunne gøre sig gældende når man bliver præsenteret for det og samtidig lave en analyse af selve udseendet af produktet? Ja det hænger jo nok lidt sammen. P: Ja, og hvilke mentale modeller det vil kunne fordré, ja.

Continued on next page

Continued from previous page

11	Recruitment is a challenge	The conditions one sets for a study isn't the challenging part, but recruiting representative users is	135- 146	F: Det jeg tænker er, at du også kommer ud i de situationer - i hvert fald med noget af det medicin der laves på Novo Nordisk og de devices der til, er at der ligesom ikke er koncentreret mængder af brugere nok til at man rejser til én stat, tester også rejser hjem igen, men det her med at du faktisk skal rundt til de forskellige nærmest. Om man kan komme rundt om det på en eller anden mere formuftig måde; ét i forhold til ressourcer, men også to, det her med du siger at konteksts aldrig vil være holdt konstant, hvis man ligesom rejser rundt til forskellige dele og nye usability forhold osv. P: Jeg tænker ikke det er forholdene der er udfordringen der, det er mere det med at rekruttere de rigtige. P.t. er jeg ved at planlægge et usability studie for en kunde, hvor vi regner med at skulle rundt til... vi regner med at kunne rekruttere fem til seks deltagere i hver større by, så vi skal rundt til en... Jeg tror vi har regnet med fem større byer i USA, til den summative test. Det er simpelthen fordi de er svære at komme i nærheden af, de har super svagt immunforsvar og det er generelt svært at få dem ud i testlokaler og der er også risiko for at de smitter hinanden og så videre.
12	Authoirty demands influences evaluation and sample	Demands from authorities decides type of sample and 'method', as certain criteria has to be fulfilled and proven right i.e. specific safety criteria	147- 152	F: Ja, men hvilke forbehold tager I jer så, i forhold til det? Altså simpelthen det her med, ét det her med at de ikke kan komme ud, men to, også at der er ret mange variable der ændrer sig, hvis I skal ud. P: Jamen altså, nogen af dem kan jo ikke komme ud og det er dem vi gerne vil rekruttere og vi ved også at det rekrutteringsbureau vi har fat i, har lavet studier med den her brugertype før. Men ellers siger de at det typisk er telefoninterviews de laver med den her brugergruppe, men nu har vi altså et device med et nyt koncept som skal testes også er der ikke rigtig nogen vej udenom, i forhold til at efterleve kravene fra myndighederne.
13	Budget increases necessary	Increasing the budget for the recruitment process is necessary, when targeting hard-to-reach populations for one's sample	153- 159	F: Tager I jer så nogle forbehold i forhold til testen? Prøver I ligesom og sige at i kender... Der vil være de her begrænsninger i forhold til det der og ligesom prøve at godtgøre på det forhånd eller? P: Ja øh... Altså vi har taget de forbehold at budgettet er større og har sat flere penge af til at rekruttere og hvis vi skal bruge altså pr. 15 testdeltager skal vi bruge 18 i tilfælde af aflysninger eller folk der ikke dukker op. Vi har sat både tid og penge af til og rejse til seks større byer i stedet for én eller to. Så hvis man skal tage yderligere forbehold, så skulle man sætte endnu flere penge og mere tid af og besøge flere byer ud fra en worst-case betragtning af at man måske kun får to testdeltagere i hver by.
14	Control all variables	As a researcher one is responsible for ensuring the same context for each participant, no matter what. This somehow influences the options one has when choosing facilities	160- 167	F: Ja, det vil ikke være så eftertragtet kan man sige. Men jeg tænker også det her i forhold til med at nu rejser i rundt til dem og omstændighederne er måske anderledes i den ene by i forhold til en anden. Er det noget i ligesom tænker ind i de resultater i nu en gang får? P: Det er jo vores opgave og sørge for at konteksten og testmiljøet, testlokalet og hvad der ellers er... Det der er relevant og kan spille ind på interaktionen har meget... At vi kontrollere dem og simulerer dem som værende ens. Der kan det stille nogle krav til testfaciliteterne og til rekrutteringsbureauet, hvis det er dem der sørger for testlokale. Der er så alligevel lidt begrænsninger, for eksempel sådan noget som gulvtæpper... De fleste af de her typer lokaler, de har desværre gulvtæpper, hvilket gør det sværere at desinficerer sådan et rum.

Continued on next page

Continued from previous page

15	Internal tests validity situational	Goal of evaluation or demands from higher authorities decides whether internal tests are sufficient	176-184	F: ... Synes det ude ved jer at være validt nok at lave de her interne tests? P: Ja, alt efter hvad vi testede og hvad vi ville undersøge, så ja. Vi benyttede blandt andet HR-medarbejder til eye-tracking studier, hvor vi også havde laboranter som repræsenterede superbrugere. Nu ved jeg ikke om det er sådan et særligt kendetegn for en HR-medarbejder, men de er ikke så bange for at tage ansigt. Det kan der selvfølgelig være andre brugergrupper der er, men man kan sige det kommer også an på om det er større designbeslutninger der skal træffes, så basere vi dem nok ikke udelukkende på tests.. altså hallway tests i gangen, der lige er udført. Der har også været nogle gange, hvor måske produktejeren eller en software chef har, simpelthen krævet at der er lavet brugertest med rigtige brugere, hvis det er at de skal lave en større designændring, som ikke er planlagt.
16	Goal of evaluation decides type of sample.	The burden of proof is another aspect in the validity of using non-representative samples. For validations end-users has to be recruited and for less important decisions it is possible to use internal employees	185-194	F: Okay. Jeg startede sådan med og have en meget teoretisk tilgang til det og læste op på sampling og populationer. Jeg synes ligesom at al litteratur siger det her med at samplen skal være helt repræsentativ, for at det du får ud det reel set kan generaliseres. Jeg hørte lidt på dig som, at det lige som kan efterkommens, hvis konteksts og de perceptuelle og kognitive faktorer der ligesom er vedligholdt. P: Njah... Ikke i en validering. Altså i forhold til hvis du er der, hvor du skal finde ud af skal vi have punktstørrelse 10 eller 12 i vores labelling eller sådan noget, jamen så kan du godt have nogen ind fra nabokontoret. F: Okay - så det er meget hvad du tester der afgører det? P: Ja og graden af bevisbryde som testen egentlig skal fremlægge. F: Ja, så hvad det skal bruges tilbagefter? P: Ja, altså om det er en validering eller det er en... noget mindre vigtigt.
17	Self-inflicted issues of not evaluating with users	Keeping a product secret and only to be evaluated with people inside a business-framework is a self-inflicted issue. It is better to get feedback from users and when a participant signs an NDA, one has to accept it as 'good enough'	195-206	F: En anden problemstilling jeg arbejder ud fra, det er det her med, hvis du har et produkt, der bliver så konfidentielt at det ikke kan vises udenfor firma-konteksts. Jeg ved ikke om det er noget der gjorde sig gældende ved Radiometer eller der hvor du sidder nu? P: Altså vi har sådan en standard non-disclosure agreement, som testdeltageren skriver under. F: Ja og så kan alle deltage når den er skrevet under? P: Ja. Altså jeg vil hellere have god feedback på min ide, end jeg vil holde den sådan fuldstændig hemmelig. F: Ja, altså idéen fremsprang fra min vejleder fra Aalborg, som er tilknyttet B&O og der har der ligesom været... Ikke en skjult agenda, men når designet lå sådan nogelund fast, så blev det meget at nu kan det ikke vises til nogen og nu er det ligesom in-house. Han havde sådan lidt en idé om hvordan det kan løses og hvordan kan det testes og stadig få valide resultater, uden og inddrage både end-users, men egentlig også almindelige users. P: Ja... altså jeg synes ikke rigtigt præmissen holder. Hvis testdeltageren underskriver en fortrolighedserklæring, så må man sige det er godt nok. Jeg synes lidt det er noget fis at man... når man fastlægger designet så ikke vil vise det til nogen. Nu ved jeg godt der selvfølgelig kan være mange stakeholders, men jeg synes lidt det er noget fis. Det er sku et selvskabt problem.

Continued on next page

Continued from previous page

18	Expert evaluations	It isn't always necessary to evaluate with users. Experts can be used for evaluating smaller design changes, or if one can argue that the change in design won't affect the use of it.	214- 222	F: Men tænker du i forhold til - hvis vi siger det er et scenarie, at der kan gøres noget i forhold til at du udelukkende ekspert-bedømmer noget eller tænker du simpelthen at end-useren skal med indover eller er det nok bare endelig og validerer? Er der et eller andet man kan gøre for at det vil være okay? P: Altså du kan godt... Der er eksempler hvor du godt bare kan evaluerer eller have en ekspert evaluering, det mener jeg bestemt. Der hvor jeg typisk har brugt det, er ved mindre designændringer, hvor du for eksempel kan sige at vi har evaluert brugen af det her produkt og nu ændrer vi kun den her ene del, for eksempel lad os sige bagcoveret skal nu hvor mørkeblåt i stedet for bordeaux rødt, så kan du lave en ekspert evaluering af det, ikke? Men mit fokus har meget været på det her med sikkerheden, altså der er nogen designændringer der let kan argumenteres for ikke at påvirke den måde hvor programmet betjenes på.
19	Substitution of population impossible in some situations	Occasionally a targeted population will have experiences and/or knowledge that isn't possible to obtain in a substitute sample.	225- 242	F: Ja, altså man kan sige det jeg arbejder frem imod, er hvordan jeg skal gøre det an, hvis den egentlige population man designer til, ikke er tilgængelig; enten fordi man er limiteret af ressourcer eller fordi det simpelthen bare ikke kan lade sig gøre - men jeg synes måske... Det er den samme problemstilling jeg ender i, igen og igen... Så snart at det ligesom er noget meget håndgribeligt, eksempelvis en lille designændringer eller det er "Kan du finde ud af at skifte batterier i det her" så du'r alle brugere, fordi de kriterier der ligesom skal opfyldes, de kan opfyldes af alle. Men så snart du ligesom kommer til det punkt at det skal være en slutbruger eller i hvert fald en bruger med en hvid viden, så kan du ikke bare fodre dem en kontekts. Jeg havde eksempelvis det problem at, hvis man lod som om det var diabetikere man ikke kunne få fat i, så bare det at fortælle dem at de ville skulle stikke sig 3-5 gange om dagen, det virker af meget for en almindelig person, men for en diabetiker er det fuldstændig acceptabelt, for det er de vant til. Så hvordan man ligesom får fodret noget kontekts ind, hvor det mangler uden at man trækker data skæv og sådan noget. Jeg ved ikke om du har et bud på det eller om du mener det ikke kan lade sig gøre? P: Uden sådan lige og have tænkt særlig meget over det, så hælder jeg til at situationer som den der, dem kan du ikke simulere. Du kan ikke tage nogle naive eller andre brugere som testdeltagere. Det tror jeg simpelthen ikke. Altså ligesom... Jeg tror heller ikke, hvis du tester høreapparater, tage hørerne også at de skal forestille sig at de er døve eller svaghørende. Jeg tror heller ikke helt man kan forestille sig at være stomi-patient uden øhm.... F: ... Ja at have stomi? P: Ja
20	Manipulation and simulation can be used in difficult situations	Some situations can be simulated, due to ethical or other problematic aspects, but it is however hard to manipulate feelings, opinions and knowledge.	243- 255	F: Nu er det meget ekstremer, men simpelthen det her med at der er altså en grænse for, hvor meget kontekts man kan fodre, uden at det vil skævvride resultater? P: Ja, altså i forhold til at du nu siger kontekts... jeg tænker sådan... F: Ja, eller erfaring-agtigt eller viden... P: Ja, altså viden, følelser og holdninger er jo svære at putte ind i folk. Men du kan godt... hvad skal man sige... Hvis det er en Epinephrine pen, kan du godt, i hvert fald for myndighederne, få lov at simulere stress på en anden måde end at... F: Ja, end at finde folk med rigtigt stress? P: Det er mere det der med... Du ved, hvis du har set Pulp Fiction eller et eller andet andet, det er liv eller død, hvis du ikke injicerer den her sprøjte ind i dig selv, jamen så dør du. Og hvis du skal teste det på børn, så er der også noget etisk i forhold til det og hvordan siger du det til et barn og forstår de overhovedet det? så der er andre måder at simulere et tilsvarende stress-niveau på. Det kan godt ske konteksten ikke er den samme, men der er andre muligheder for at simulere et tilsvarende eller passende stressniveau på.

Continued on next page

Continued from previous page

21	Substitute samples has to be considered	In situations where a substitute sample is applicable, one might not substitute anyways. The end-users will provide richer data useful for the design-process. This is both true for summative and formative evaluations	262- 271	F: Jeg arbejder som sagt henimod det her med, om man kan substituerer den her population, men som du nævner, i de fleste tilfælde, så er det bedre at smide de ekstra penge efter en bedre rekruttering. Tænker du det for kun de summative altså de endelige tests, eller? P: Nej, det tænker jeg netop ikke. Ikke for allei formative tests, men for mange formative tests også. En ting er at du kan... Hvis vi tager det her eksempel med fontstørrelser og at du så får Hr. og Fru Jensen ind til at bedømme noget med fontstørrelser, så får du kun evalueret eller vurderet på det. Hvis du har rigtige brugere inde, jamen så får du også noget rigere data og noget yderligere input, som du egentlig ikke har spurgt om, men som kan vise sig at være virkelig værdifuldt til at informere din designprocess. Så det er også vigtigt ikke bare i de summative, for der kan du ikke påvirke dit design særlig meget, hvis overhovedet, men at få det ind i de formative faser, hvor du rent faktisk har et design der er modtageligt for input, i forhold til og ændrer lidt. F: Ja, så hvad skal man sige... At den egentlig målgruppe rent faktisk giver mere end man først regner med? P: Ja, den giver mange gange mere end bare det du mäter på og spørger om.
22	Additional data from end-users	End-users provides more than what one are measuring or asking for.	272- 273	

Chapter F

Detailed Description of Methods

In this appendix the methods included for the process of identifying a substitute sample is described. This includes a detailed description of *Personas* (and the discount-method identity Models) the *Hierarchical Task analysis* and the *Systematic Human Error Reduction and Prediction Approach*. The procedure of using the methods is described, in order to give an understanding of how to use the two methods. Additionally, for each method an example of the output is shown. This is done in order to show how the information gathered through the analyses are used.

Personas

Designing products for users often means talking to the users and determining their needs. As development of products happens in agile environments, where everything can turn around suddenly, it can be difficult to always have users to evaluate on decisions and changes to the product.

The use of Personas

Instead of talking to users every time a change is needed, personas can be helpful. Personas are used in situations where it is wanted to successfully accommodate a variety of user needs. Taking care of multiple types of needs, is difficult and accommodating for everything to obtain a product likeable by everyone isn't always possible. Figure F.1 exemplifies why it isn't possible to condense all needs into a single design (Cooper et al., 2007).



Figure F.1: Example on why putting all user needs into one design might ruin the experience and functionality for everyone (Cooper et al., 2007).

For the personas to function as intended, the right individuals have to be identified. The right individuals are those who have needs that generally represents needs from larger parts of the population. While user needs are important, so is motivations. The motivation is important when designing personas, as this is what drives the user. Motivation helps to understand *why* a user has the specific needs but also reveals *why* and *how* the user acts in different situations (Cooper et al., 2007).

Cooper et al. (2007) is the first to mention personas, and the description here, is solely based on his research in personas. He explicitly explain that personas can help overcome multiple issues and that it is a powerful, multipurpose tool (Cooper et al., 2007). Without going fully into detail, he lists the five strengths of personas, when using them.

- Determination

Through personas it is possible to determine how a product should function and decide why that is.

- Communicate

Personas can be used for argumentation and communication with all stakeholders of the project. It can determine a common goal and help the design to always be user centered, by revisiting the persona often.

- Build Consensus and Commitment

The narrative a persona always have creates a common language for stakeholders and developers. It is easier to relate to users needs when told as a narrative, instead of complex graphs and models.

- Measure

Personas can be used for testing. It doesn't eliminate the need of actual user testing, but playing out the persona can give powerful feedback. It provides a reality-check for the developers.

- Contribute

Personas can contribute in many other ways. It is possible to draw parallels into marketing and sales plans.

How to Develop Personas

The construction of personas can be done through a seven step process.

1. Identify behavioral variables
2. Map interview subjects to behavioral variables
3. Identify significant behavior patterns
4. Synthesize characteristics and relevant goals
5. Check for redundancy and completeness
6. Expand description of attributes and behaviors
7. Designate persona types

Identify behavioral variables

Identifying the behavioral variables is done from data. The data has to be about the target population, as it is ought to determine their attributes and characteristics. One type of data could be observational data. Observing user and listening to what they actually articulate about certain topics, enables one to list some behavioral variables. Typically, there's five types of behavioral variables:

- **Activities** - What the user does; frequency and volume.
- **Attitudes** - How the user thinks about the product domain and technology.
- **Aptitudes** - What education and training the user has; capability to learn.
- **Motivations** - Why the user is engaged in the product domain.
- **Skills** - User capabilities related to the product domain and technology.

Map interview subjects to behavioral variables

After having noted all behavioral variables all interview subjects should be mapped to each variable. Some of the variables might be continuous, while others are discrete choices. The precision of the mapping doesn't matter much. It just have to be visually possible to see differences, as these differences are going to play a larger part, when identifying patterns. The easiest way of explaining this step of the process is by visualizing it. In Figure F.2 an example of how to map users to the variables are shown.

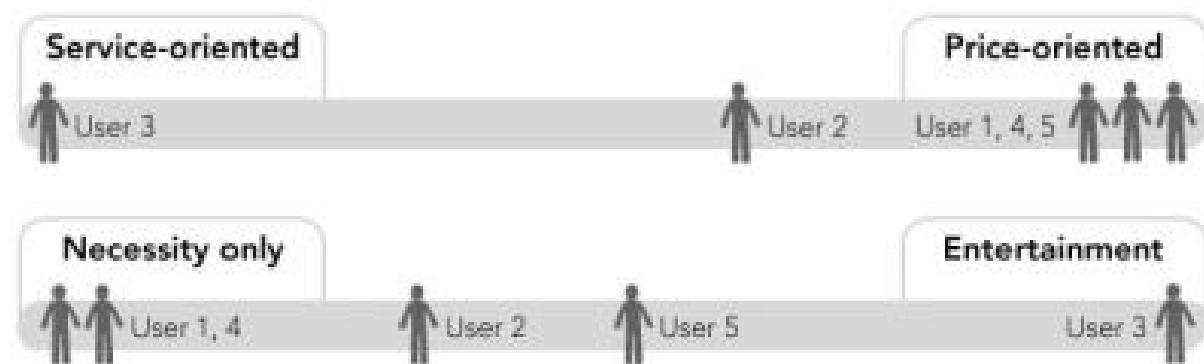


Figure F.2: A depiction of how actual users are mapped to the behavioral variables. This procedure allows for identifying patterns e.g. a cluster of users being mapped similarly across variables.

Identify significant behavior patterns

After having mapped all subjects to the behavioral variables, it should be possible to identify patterns. Cooper et al. (2007) states that a pattern typically is a set of subjects that cluster on six to eight behavioral variables. If this is the case, there evidence for a more general characteristic of the target population. For a pattern to be valid, there has to be a logical connection between the variables. A logical connection between variables often means that they support each other or are coherent - it can't just be a normal correlation.

A logical connection could be users who are price-oriented also focus on necessity only. A correlation on the other hand, could be that all user who are quality-oriented also happens to be good with animals. A correlation might be proved between the two, but there's no logical reasoning behind it.

Synthesize characteristics and relevant goals

Having identified each behavioral pattern enables one to synthesize the details of these. It is the development of a bullet list with descriptions of the behavior pattern; potential use environment, typical workday, current solutions, frustrations and relationships. There are many aspects to consider, but in this step a short bullet list is adequate.

The most important aspect of these descriptions, is that they are based on observations. If fictional reasoning gets involved, it lowers the credibility of the finished persona. The goals of a persona is the most important thing to synthesize from.

It is possible to infer the goals that leads to certain behaviors, by using the logical connection of the behavioral variables. The goals are often obtained through observation actions and analyzing participants answers to goal-oriented questions (Cooper et al., 2007).

Check for redundancy and completeness

Having identified multiple personas and having defined the goals for each of these allows for a check of redundancy and completeness.

By comparing the mappings and personas' characteristics and goals, it is possible to see if there are anything missing, that needs to be done. Further, if two personas are alike, two options remain; tweak the characteristics of one of them (in a degree where it still fits observations and data) or eliminate one.

Having no redundancy and making sure that the personas are complete is what characterizes good personas. Personas should represent the diversity of the target population and has to include the same diversity of behavior and needs.

Expand description of attributes and behaviors

From the bullet list just developed, an extensive description is now to be developed. This narrative should be no longer than one or two pages of prose. The narrative has to be written in third-person, as this way of telling is much better at conveying the persona's attitudes, needs and problems.

Designate persona types

Having multiple personas with different needs and goals, requires the researcher to designate the persona types. It is a prioritized list, but not fully. The primary persona will be prioritized, but it is ought to implement the other personas' need, without interfering with the needs of the primary persona.

When designating, there are typically six types:

- Primary

Represents the primary target for design decisions. There is only one primary persona for each project.

- Secondary

Represents the group of users that will be mostly satisfied with the needs and goals that the primary persona states. There can be a single need from the secondary person,

which can be implemented, if and only if, it doesn't interfere with the needs and goals of the primary persona.

- Supplemental

Supplemental personas are personas whose needs and goals are completely covered by any combination of primary and secondary personas.

- Customer

Customer personas focuses on the needs of customers. Not needs of end users. The customer may very well have needs identical to the primary or secondary personas.

- Served

Served personas isn't exactly end-users, but are instead directly affected by the use of the product. A patient being x-rayed is not the actual user of the system, the nurse is. But the patient might still be served very well, by a good design of interface. The served persona shows second-order social and physical ramifications of product use.

- Negative

Negative personas are typically used for communicating who the design and product are **not** being designed for. Their use is purely rhetorical: to increase the understanding of who the product has as a target.

Identity Models

Identity models are the discount version of personas. They are built through a similar process as personas, but doesn't require the same amount of resources or efforts. It starts by collecting data on people from the target population. The data collected can be anything from behaviors, identity elements, personal values etc. All data regarding the user is captured through interviews and are put into potential identity elements. Identity elements can be sources of pride, self-esteem or any other category with a value relevant to a target activity (Holtzblatt and Beyer, 2014, p. 32).

Any identity element names captured during interviews and interpretation sessions become the starting point for grouping the observations meaningfully to represent different aspects. Finally, the team of researchers agrees on the most important findings of these identity elements. The basis of picking is to pick the best stories to express the feel and identity of the user to the rest of the project team (Holtzblatt and Beyer, 2014, p. 33). As with personas, the identity model is created on the background of these identity element. The identity elements are given a short description. To give an idea of, how a finished identity model look, an example is given in Figure F.3.

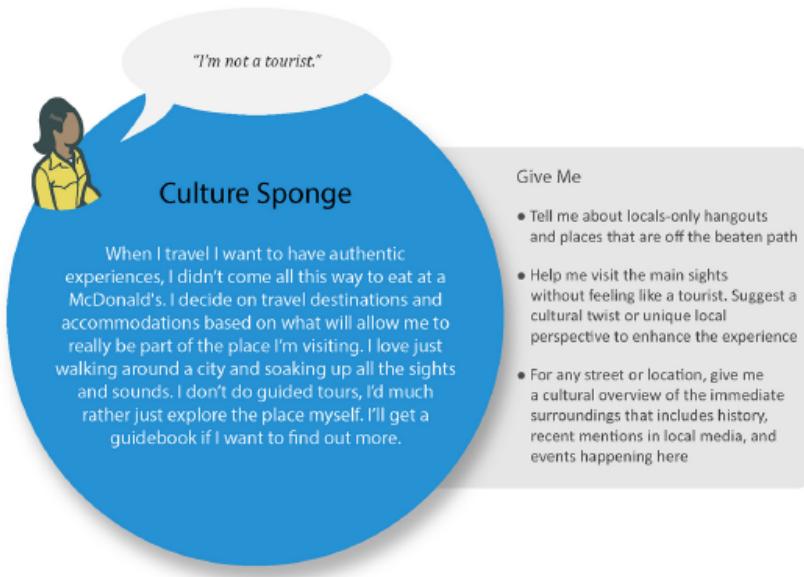


Figure F.3: Typical example of an Identity Model. It contains a description of the identity of the user. Additionally it provides a box with "Give me", which is the needs and goals of this identity.

The differences between the persona and the identity model, really is that amount of effort put into it. The identity model can be further coupled with other models, in order to show the bigger picture. This however quickly becomes as resource intensive as developing personas.

Hierarchical Task Analysis

Stanton et al. (2017) states that *Hierarchical Task Analysis* is the most popular task analysis method. The outcome of the method is an exhaustive description of any task activity from the specific evaluation which provides detailed insights to the demands of each individual task. These insights are used for greater understanding of the evaluation, which is achieved through the following six step procedure:

Defining tasks under analysis

The tasks of interest has to be identified, as these are the objects to be analyzed. Simultaneously, the general purpose of doing the HTA should be defined in the beginning. This enables one to always revisit the goal and purpose of the analysis, if uncertainty or ambiguity occurs in any of the following steps.

Data Collection Process

The goal of this step is to inform the development of the HTA. This includes collecting data on the tasks involved. This could be aspects as the kind of technology used, the required human-product interaction, decision making and foreseeable task constraints. The method for collecting this type of data isn't predefined, but instead multiple method can be used. Such method could rely on observations, interviews, questionnaires or walk-throughs. The method of choice comes down to the effort of the analysis and natural constraints i.e. time, resources and access.

When the two first steps of the procedure have been taken, the actual development of the HTA begins.

Determine the overall goal of the tasks

The overall goal is the desired and intended action which is required for the specific task. The task could be "Please use this device to take a dose of 20 units. When you are certain that you have done the task, please indicate this by placing the device on the table in front of you". The overall goal of this would be 'Inject 20 units and place the device in front'. The simplicity of the step shouldn't be perceived as wrong or too vague - the next steps will go further into detail.

Determine the sub-goals of the task

This step breaks down the overall goal of the task into smaller tasks needed for achieving the goal. Reusing the example from before, this could imply sub-goals as 'Set the correct dose'.

The only rule for this step is that the combination of the sub-goals, leads to the overall goal. This means that all sub-goals are mandatory to fulfill, in order to achieve the overall goal.

In some situations there could be sub-goals that isn't fulfilled, but still the overall goal is achieved e.g., a sub-goal demanding one to use alcohol swipes on the injection site, before injecting. While such goal is skip-able, it doesn't directly hinder a participant achieving the overall goal. Such situations are very dependent on the goal of the evaluation, meaning if it's important for the evaluation, these observations of unfulfilled sub-goals should be noted down.

Sub-goal Decomposition

When decomposing the sub-goals, each sub-goal might be broken into multiple sub-goals. This steps break down all sub-goals into actual operations a participant has to do for achieving the associated goal. There no correct amount of sub-goals, when decomposing, the only rule is that the process is used, until meaningful operations are identified. Having all operations for a task, makes it possible to proceed to the final step of the task analysis.

Plans analysis

Having completed all steps before this, enables one to add plans for each task. The plan describes the order in which each operations needs to be done achieve the overall goal of the task.

Outcome of HTA

The outcome from following these six steps can be given as a flowchart or in a tabular format. It defines the operations a participant has to do and the order thereof. This allows for insights of each task and the sub-goals. Investigating the steps can help one identify possible areas where issues will occur for the participant, hence showing where usability problems might occur. Stanton et al. (2017) have already exemplified how the outcome from an HTA could look, which is depicted in figures F.4 and F.5 on the next page.

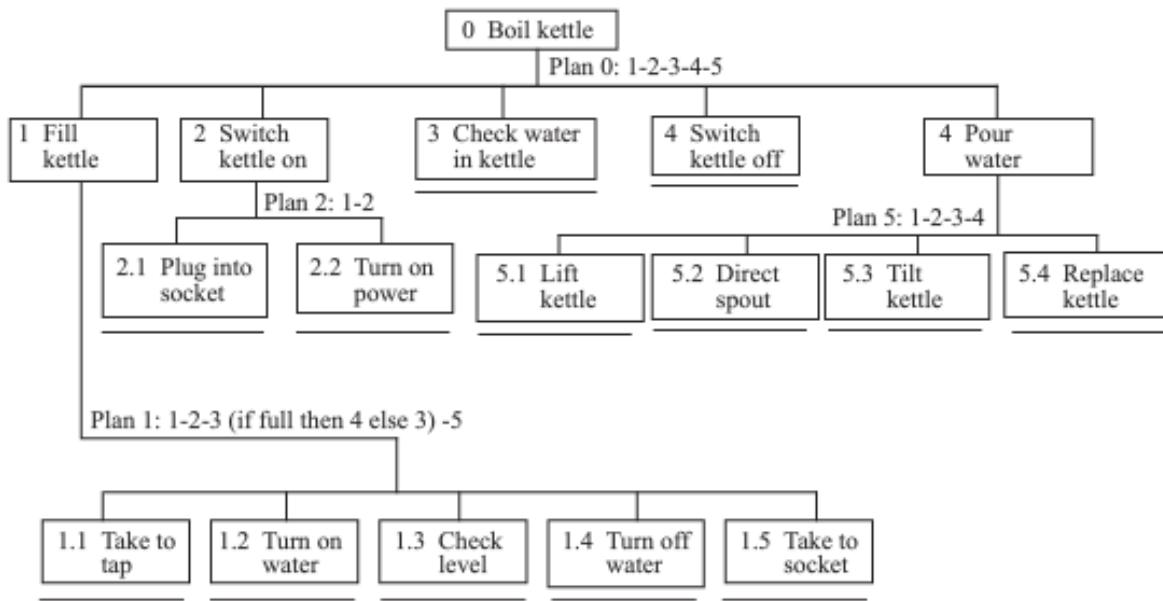


Figure F.4: Depiction of HTA Flowchart for the task "Boil water". It shows how the overall goal can be further split in to sub-goals and operations. It also shows how plans can enable the understanding of sequence.

0. Boil kettle	
Plan 0: Do 1 then 2 then 3 then 4 then 5	
1. Fill kettle	Plan 1: Do 1 then 2 then 3 (if full then 4 else 3) then 5
	Take to tap
	Turn on water
	Check level
	Turn off water
	Take to socket
2. Switch kettle on	Plan 2: Do 1 then 2
	2.1 Plug into socket
	2.2 Turn on power
3. Check water in kettle	
4. Switch kettle off	
5. Pour water	Plan 5: Do 1 then 2 then 3
	5.1 Lift kettle
	5.2 Direct spout
	5.3 Tilt kettle
	5.4 Replace kettle

Figure F.5: Depiction of HTA Table for the task "Boil water". It shows how the overall goal can be further split in to sub-goals and operations. It also shows how plans can enable the understanding of sequence.

SHERPA

The application of the *Systematic Human Error Reduction and Prediction Approach* is generic and is applicable in all domains involving human activity (Stanton et al., 2017, p. 143). The method consists of an error mode taxonomy and is further linked to a behavioral taxonomy. These taxonomies are applied to the aforementioned HTA with the goal of predicting errors induced by either design or humans. Stanton et al. (2017) states that this method is one of the most successful in terms of accuracy of error prediction (Stanton et al., 2017, p. 143).

Identical to the HTA the procedure of SHERPA is divided into steps. There's a total of seven steps in SHERPA. Successfully completing the seven steps provides insights into the evaluation and the tasks. These insights mainly revolves around the prediction of potential errors. In order to understand the method the seven steps are further described below.

Hierarchical Task Analysis

SHERPA can be perceived as an expansion of the HTA. Due to this fact, it requires a fully completed HTA where operations for each sub-goal have been identified. The goal of SHERPA is to indicate potential errors for each task of the HTA. This is achieved by using pre-established error taxonomy. The taxonomies are applied in some of the following steps.

Task Classification

In the classification of tasks a behavior taxonomy is applied. This is done by taking the first (or next) bottom level task from the HTA (e.g., the last sub-goal or the operation itself) and classify it according to the aforementioned taxonomy. The established taxonomy for behavior is:

- Action (e.g., pressing a button, pulling a switch, opening a door)
- Retrieval (e.g., getting information from a screen or manual)
- Checking (e.g., conducting a procedural check)
- Selection e.g., choosing one alternative over another)
- Information Communication (e.g., talking to another party).

When all tasks have been classified this step is complete. Next step of the method is to apply another taxonomy.

Human Error Identification (HEI)

The goal of this step is to determine the credible error modes for the tasks. This is done by applying researchers domain expertise and by using an error mode taxonomy. The error taxonomy is given in Table F.1 below.

Action errors		Checking errors
A1	Operation too long/short	C1 Check omitted
A2	Operation mistimed	C2 Check incomplete
A3	Operation in wrong direction	C3 Right check on wrong object
A4	Operation too little/much	C4 Wrong check on right object
A5	Misalign	C5 Check mistimed
A6	Right operation on wrong object	C6 Wrong check on wrong object
A7	Wrong operation on right object	
A8	Operation omitted	
A9	Operation incomplete	
A10	Wrong operation on wrong object	
Retrieval errors		Selected errors
R1	Information not obtained	S1 Selection omitted
R2	Wrong information obtained	S2 Wrong selection made
R3	Information retrieval incomplete	
Communication errors		
I1	Information not communicated	
I2	Wrong information communicated	
I3	Information communication	

Table F.1: The error mode taxonomy established in the SHERPA. Each error is referred to by its number e.g., 'operation too long/short' is assigned A1.

When each task have been assessed with respect to credible errors, a short description of the specific credible problem is given. If the credible error is C1 (Check omitted) a description should suggest how and why.

Consequence Analysis

Knowing what kind of credible errors that could happen and having a descriptive reason, makes it possible to determine the consequence of this particular error. In this step of SHERPA the goal is therefore to determine and describe the consequences associated with the predicted errors identified from the step before. Each description should explain the consequence in relation to the associated task.

Recovery Analysis

Knowing potential errors and the consequences enables one to check for recovery. Determining the recovery potential of an error can be difficult and there's also situations where recovery simply isn't possible. Such situations are assigned 'None' in this step.

An important aspect in this step, is that if there's a later step in the task that enables a potential recovery, this should be noted here e.g., a constraint preventing a certain interaction forcing the participant to revisit and solve the error.

Ordinal probability Analysis

With the knowledge obtained through the first couple of steps, it should be possible to rate the probability of an error happening. Knowing the type of error, the consequence and the recovery potential leads to a well-informed assessment of probability. Most commonly a ordinal probability is used, thus assigning each error with low, medium or high. Stanton et al. (2017) sets a couple of rules for assessing the probability, which are:

- if the error has not occurred previously then a low (L) probability is assigned.
- if the error has occurred on previous occasions then a medium (M) probability is assigned.
- if the error has occurred on frequent occasions, a high (H) probability is assigned.

Criticality Analysis

As the last step of the SHERPA, the criticality of each error is assessed. The same ordinal scale used for assessing probability is used, but the way of thinking is changed a little. If an error would lead to a critical incident (product or human-wise) it should be rated as high (H). Completing all steps until now should help one obtain a completed SHERPA.

Outcome from SHERPA

While the seven steps have provided a structure for how to use the method, there's still not given any format for outcome. Typically the outcome from a SHERPA is a table. The table contains all the information from the steps and gives an overview of the evaluation and the tasks it consists of. In figure F.6 on the next page an example of SHERPA output is depicted. The example is from Stanton et al. (2017).

Task Step	Error Mode	Error Description	Consequence	Recovery	P	C	Remedial Strategy
1.1	A8	Fail to switch VCR on	Cannot proceed	Immediate	L		Press of any button to switch VCR on
1.2	C1 C2	Omit to check clock Incomplete check	VCR Clock time may be incorrect	None	L	!	Automatic clock setting and adjust via radio transmitter
1.3	A3 A8	Insert cassette wrong way around Fail to insert cassette	Damage to VCR Cannot record	Immediate Task 3	L L	!	Strengthen mechanism On-screen prompt
2	A8	Fail to pull down front cover	Cannot proceed	Immediate	L		Remove cover to programming
3.1	S1	Fail move timer selector	Cannot proceed	Immediate	L		Separate timer selector from programming function
3.2	A8	Fail to press PROGRAM	Cannot proceed	Immediate	L		Remove this task step from sequence
3.3	A8	Fail to press ON button	Cannot proceed	Immediate	L		Label button START TIME
4.1.1	A8	Fail to press UP button	Wrong channel selected	None	M	!	Enter channel number directly from keypad
4.1.2	A8	Fail to press DOWN button	Wrong channel selected	None	M	!	Enter channel number directly from keypad
4.2	A8	Fail to press DAY button	Wrong day selected	None	M	!	Present day via a calendar
4.3	I1 I2	No time entered Wrong time entered	No programme recorded Wrong programme recorded	None None	L L	!	Dial time in via analogue clock Dial time in via analogue clock
4.4	A1	Fail to wait	Start time not set	Task 4.5	L		Remove need to wait
4.5	A8	Fail to press OFF button	Cannot set finish time				Label button FINISH TIME
4.6	I1 I2	No time entered Wrong time entered	No programme recorded Wrong programme recorded	None None	L L	!	Dial time in via analogue clock Dial time in via analogue clock
4.7	A8	Fail to set timer	No programme recorded	None	L	!	Separate timer selector from programming function
4.8	A8	Fail to press TIME RECORD button	No programme recorded	None	L	!	Remove this task step from sequence
5	A8	Fail to lift up front cover	Cover left down	Immediate	L		Remove cover to programming

Figure F.6: The typical format for the outcome of a SHERPA. It breaks down all tasks and potential credible errors and provides strategies for remedying or recovering from potential error.

Chapter G

Diabetes in General

What is Diabetes?

Diabetes is a collection of disorders which together makes the human body resistant or incapable to produce insulin. The disease makes the body unable to independently produce and process insulin. Both cases hinder the needed glucose uptake, which causes hypoglycemia, thus resulting in increased blood glucose as the uptake of glucose by organs doesn't happen.

It is mainly caused by a disorder in the pancreas and the consequence of the disorder is, that the blood glucose doesn't regulate automatically Turner and Robinson (2014). In general, diabetes can be categorized as type 1 diabetes mellitus (T1D) or type 2 diabetes mellitus (T2D).

T1D is a chronic immune disease and is hereditary. It is often diagnosed in the childhood, hence people diagnosed with T1D have lived with the disease for a long period. This type of diabetes is caused by a malfunction in the immune system. The malfunction causes the immune system to classify the insulin-producing cells as diseases and attacks them. This results in a high reduction or completely stopped production of insulin, thus making the patient insulin-dependent.

T2D is the most common type of diabetes. The disease is a combination of insulin-resistance and degraded production of insulin. At first, the body becomes resistant to insulin, making the pancreas overproduce insulin, in order to compensate for the resistance. The overproduction of insulin stresses the pancreas, and further into the course of the disease the pancreas won't be able to maintain this overproduction Turner and Robinson (2014). T2D is typically caused by what would be categorized as environmental factors and lifestyles. Regarding lifestyles, factors as cholesterol level, age and BMI are all connected with a higher risk of being diagnosed with T2D. Opposite T1D, T2D is controllable to a certain degree. This does, however, require major changes to lifestyle, diet, and exercise.

The treatment of both T1D and T2D relies on controlling and maintaining a normal level of blood glucose. While T1D is treated by subcutaneous injections, T2D can be treated with oral medication, but only in situations where adherence to both diets, exercise, and medication is maintained. Otherwise, T2D is also treated by subcutaneous injections of insulin.

Diagnosed with Diabetes

Even though the treatment for diabetes in itself seems manageable, there are several aspects that are unimaginable. Each time a patient has to take a dose of insulin, there are multiple decisions to be made. Additionally, it isn't just to inject the insulin as one would think. In figure G.1 some of all these decisions and the physical activities needed for taking a dose of insulin is shown. Laagland (2017) is the creator of the figure, which was made for his Master Thesis. He identified the challenges a diabetic goes through.

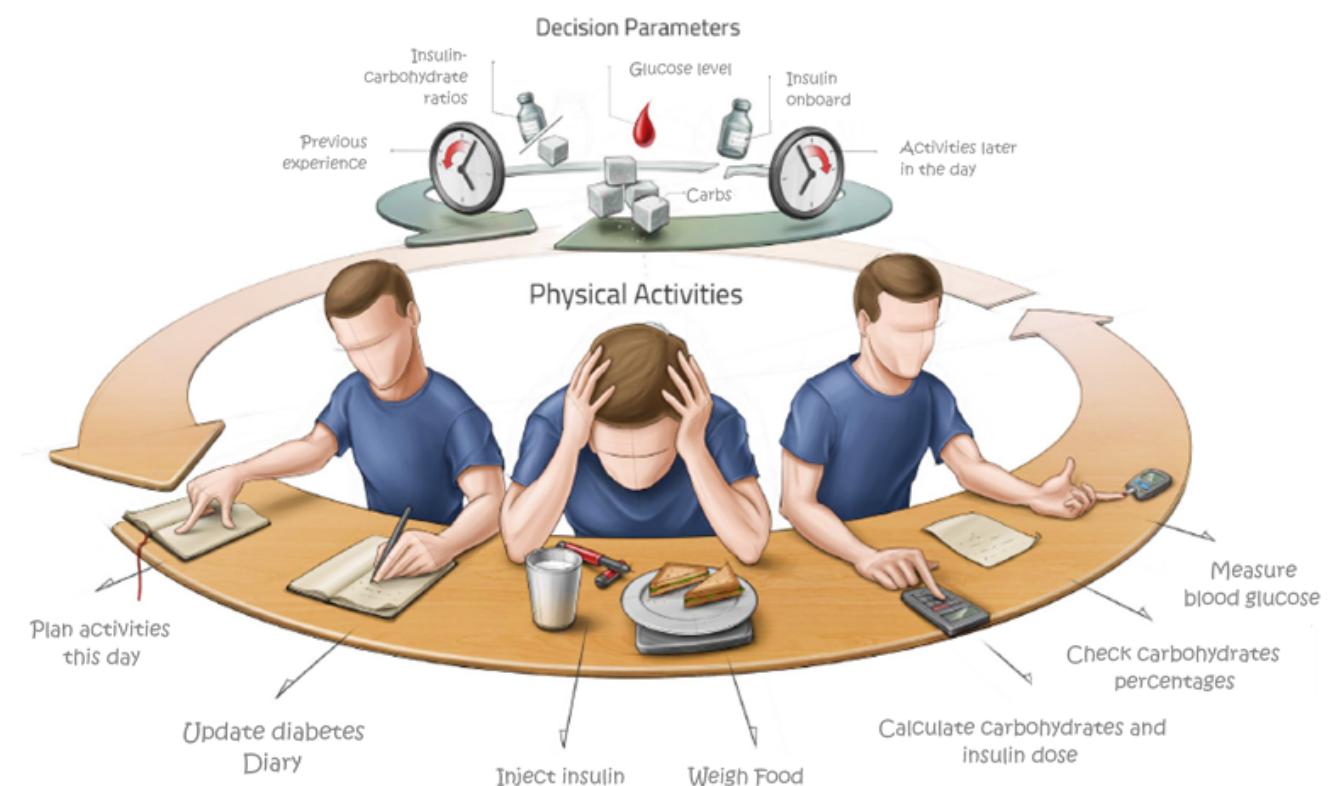


Figure G.1: A depiction of the different decisions that can influence the amount of insulin a patient has to take before each meal. Additionally a listing of all the needed physical activities that are required to do, each time. Figure developed by Laagland (2017)

As the figure depicts, a lot of stress and frustration is present whenever the medicine is needed. This is due to the multiple decisions to take into account, but also the immense work it actually takes to do the actual injection. Laagland (2017) view on living with diabetes seems coherent with other studies, where the objectives have been to map out challenges of being diagnosed and living with diabetes. The DAWN program (Diabetes Attitudes, Wishes, and Needs) (Skovlund and Peyrot, 2005; Nicolucci et al., 2013) is an effort to investigate the psychosocial and behavioral barriers that coexist with being a diabetic. It showed how diabetes-related distress was associated with poor self-management of the disease. The distress could

be associated with fear, stress, anxiety or fear due to the risks introduced by diabetes. Additionally, some described it as mentally hard to live with a disease like diabetes. Furthermore, some people with diabetes feel ashamed and misunderstood due to misinformation or uninformed people. Laagland (2017) tried to build a metaphor, depicted on G.2 on the next page, for all these factors influencing a diabetics life. It encapsulates most of the research made within the area, as it shows consists of many of the negative effects that follow having diabetes. The distress (insecurity, unmotivated and frustration) is caused by six factors which are; (1) The lack of references, (2) The lack of the right knowledge, (3) The lack of structure, (4) Health care providers only being able to give a '*leg-up*', (5) Informal caregivers have insufficient knowledge and (6) Barriers to go to others with diabetes.

Observations at Novo Nordisk

Having participated in several evaluations with samples from the targeted population has given some insights, not possible to obtain otherwise. T1D patients seem to have a better adherence to their medication and the use of it. They are simply better at controlling their blood sugar and overall better at self-managing their diabetes when compared to patients with T2D. One possible explanation for this could be that T1D is often diagnosed at early ages, thus diabetes being a part of their whole life. T2D patients, on the other hand, differentiates much more. Some are able to restructure their lifestyle and control their disease, while others focus on completely other aspects.



Figure G.2: A metaphor for the struggles a diabetic experiences. The metaphor consists of six different barriers that a diabetic experiences. It shows how these six factors makes the patient insecure, unmotivated and frustrated about the disease. Figure developed by Laagland (2017)

Interview with Anthropologist

The former student from Aalborg university Rasmus Jensen (Jensen, 2017) conducted an interview with an anthropologist employed at Novo Nordisk. Some of the findings from the interview state, that the goals of a diabetic, might not be coherent with the treatment offered by the medical devices. Instead of focusing on being healthy and trying to lose weight, the goal of a diabetic could be to maintain the feeling of togetherness and belongingness (not being different and being able to do what everybody else can do). Furthermore, the priorities of diabetics might not be what is first expected. Staying healthy, controlling the disease and being adherent to the treatment, might from our point of view seem like what needs to be prioritized. It was, however, through the interview found that providing a roof for the family, making sure everybody had what they needed could be prioritized higher than the well-being of the patient.

From the interview, Jensen (2017) created a list of 13 items describing what user experience actually was for this population. The items are given below. These items won't enter the user profile of diabetics but instead shows why the targeted population really is necessary in some cases. Further, it shows how substitute sampling is near impossible when the goal of the evaluation is to *measure* the user experience.

- | | |
|--|---|
| <ol style="list-style-type: none">1. Users need to know whether or not they have taken their medicine, which is a problem to some. As it becomes a daily routine, whether or not it has been done can be hard to remember. A little like turning off lights or locking the door, except there is no clear evidence whether it has been done before the body starts reacting.2. The experience of using a device needs to feel simple and introduce only a few things to understand. The action of taking the medicine shouldn't be a problem, as there are plenty of problems already, having diseases needing injection treatment.3. Users need to feel able to do the things they like and avoid changing into a life that makes them unhappy. The lives they live is what makes users happy and | <p>diabetes is what forces them to change into something they do not like.</p> <ol style="list-style-type: none">4. Users need to understand what is happening. If the users do not understand they will feel like a failure.5. The experience should not feel obtrusive or like '<i>it fills a lot</i>' in the users' life.6. Users do not want to feel stigmatized by their situation. Rather, the experience should feel like it allows them to stay part of the community they like.7. Users wants to feel served and relaxed.8. The experience should feel like it is easily integrated into a user's life, as users would benefit from integrating diabetes into their lives instead of pushing it away.9. Users should not feel like they need to prioritize these above priorities like e.g. feel togetherness and belongingness, as |
|--|---|

- they will not do so.
10. Users need to feel like the experience is made for them, and not for someone else.
11. Users need more than just getting told what to do, as this will not make them do it.
12. Users need to be motivated by other things than staying healthy and weight loss to actually stay in control, as healthiness and weight loss are not a goal for the user in itself
13. Users need to feel like their actions are meaningful.

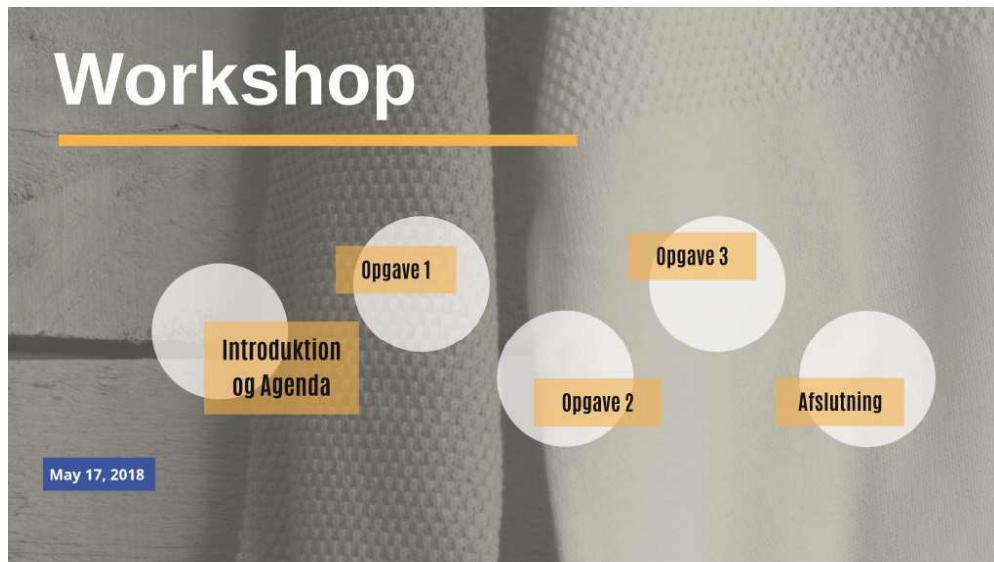
Even though these aspects of user experience won't enter the user profile of diabetics, there are still some of the items, that has a chance of being general for more than just this population e.g. item 4, as incomprehensible situations rarely provide a good user experience.

Chapter H

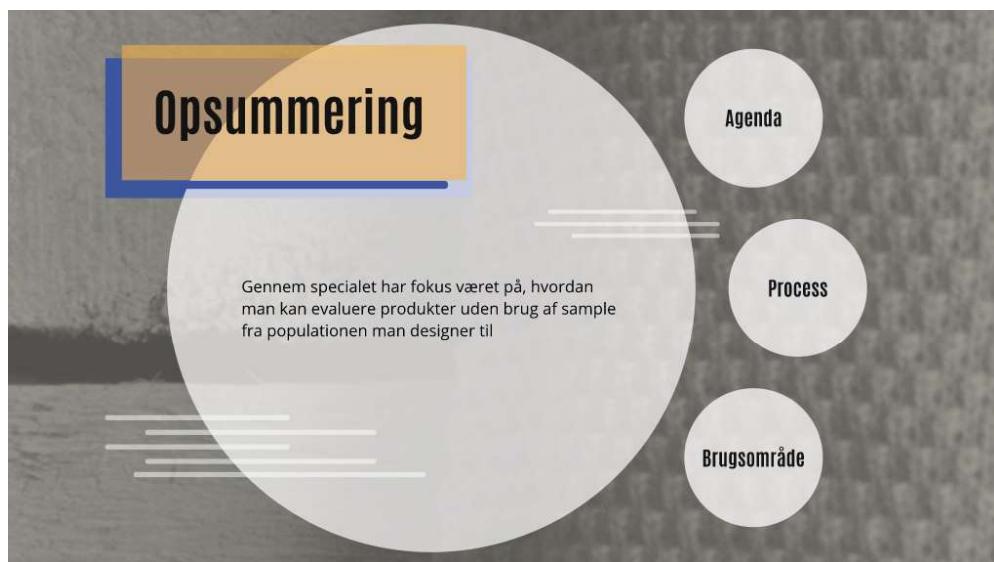
The Workshop and The Tools for Conduction

This appendix contains the slides used for presentations throughout the workshop. There's a total of 46 slides, which shows everything showed during the workshop. They do however lack the presenters notes, which was used to support the facilitation. The presenters notes are placed immediately after the slides. They are presented in a table, where the number of the note corresponds to the number of slide. Slide 6, 22, 31 and, 46 doesn't have any presenters notes, due to the slides being the fill, in-between assignments. Slides that contain animations have the same presenters notes. This causes some of the identifier to span over more than one slide e.g., slide 12-13.

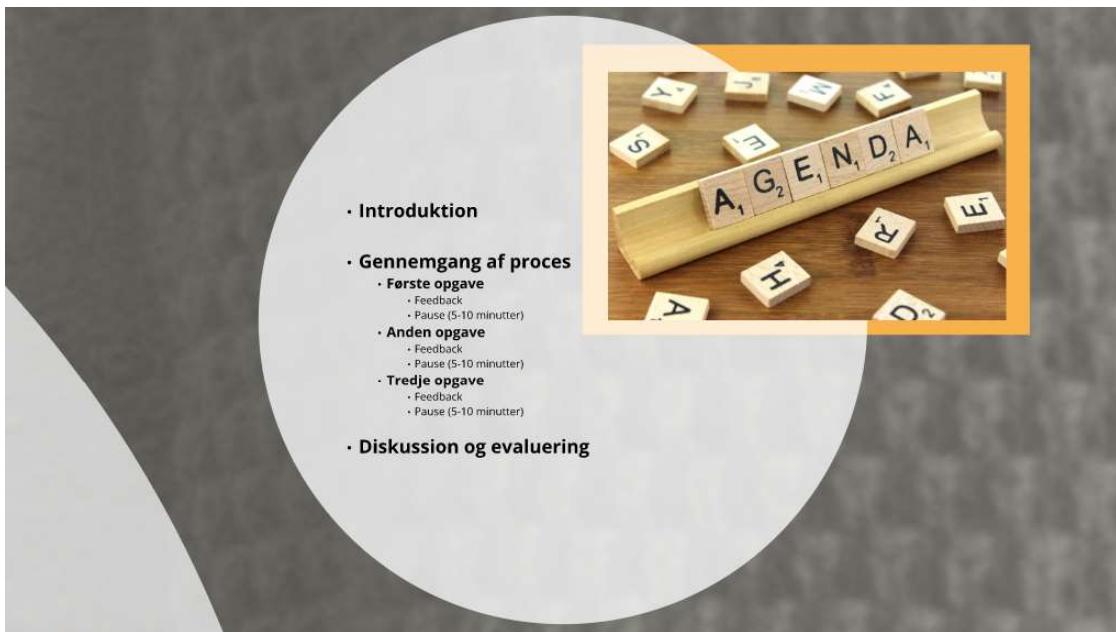
Presentation for Workshop



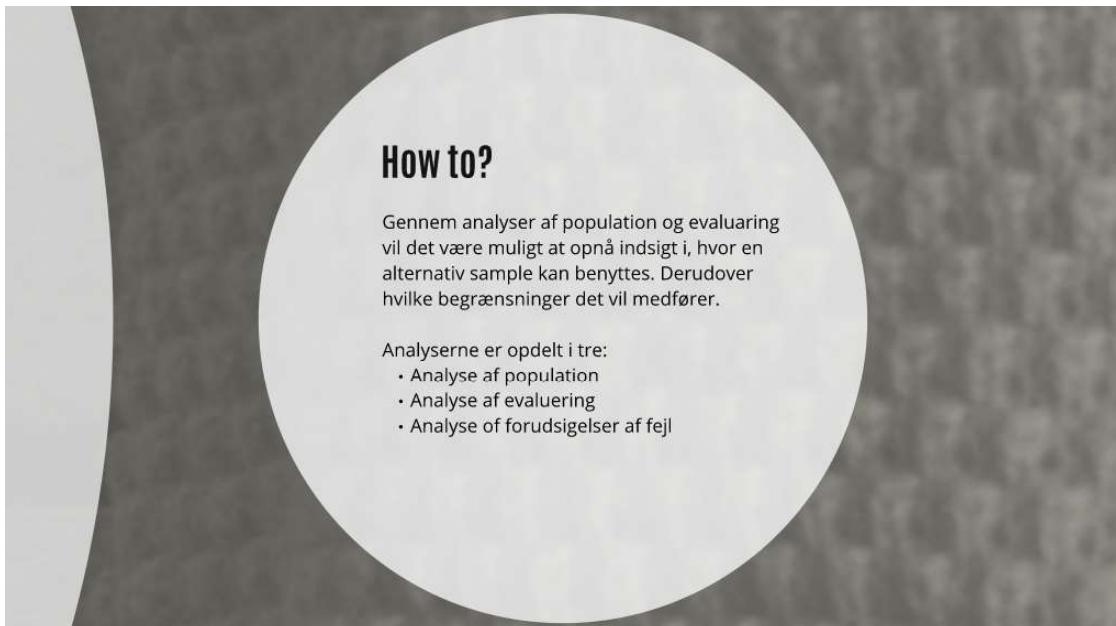
1.



2.



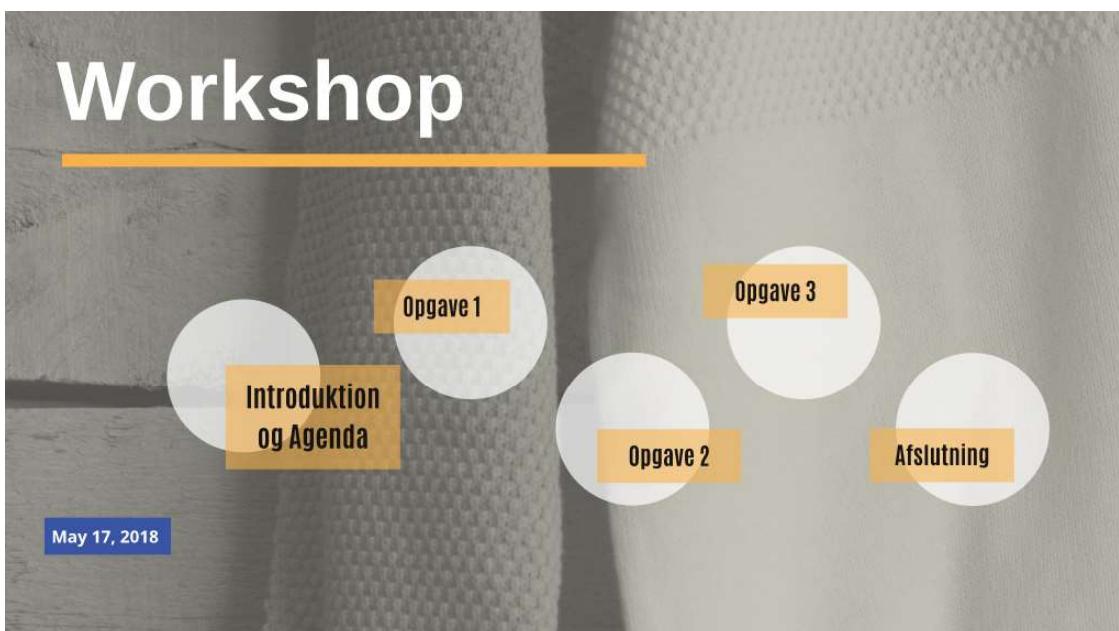
3.



4.



5.



6.



7.



8.



9.



10.

Attributter og Karakteristika

11.

Opstilling af Atributter - Step 1

Fokus spørgsmål

Hvilke attributter og karakteristika beskriver en diabetiker og gør sig gældende under en hvilken som helst evaluering?

1. Brainstorm og noter individuelt så mange adfærdsmæssige variable (atributter og karakteristika) der gør sig gældende for en diabetiker

- Ingen snak med de andre

12.

Opstilling af Atributter - Step 1

Fokus spørgsmål

Hvilke attributter og karakteristika beskriver en diabetiker og gør sig gældende under en hvilken som helst evaluering?

Aktivitet

Hvad gør brugeren og hvor tit?

Attitude

Hvad tænker brugeren om produkt-domænet og teknologien?

Aptitude

Hvilken uddannelse og træning har brugeren samt forudsætninger for yderligere læring?

Motivation

Hvorfor er brugeren engageret i produkt-domænet?

Færdigheder

Hvilke færdigheder har brugeren i relation til produkt-domænet og teknologien?

13.

Opstilling af Attributter - Step 2

2. Tag alle Post It's og placer dem på whiteboardet under den kategori de tilhører (fx. "Aktivitet")

- Hvis man bliver inspireret af andres Post It's, lav da en ny og tilføj den
- Stadig ingen snak eller diskussion

14.

Opstilling af Attributter - Step 3

3. Alle grupperer Post It's med lignende indhold under hver af de fem overkategorier.
Det er ikke muligt at kategorisere Post It's på tværs af overkategorierne.

- Flyt gerne rundt på andres grupperinger, hvis de ikke giver mening for dig
- Under grupperingen er diskussion eller snak ikke tilladt
- Step 3 er først færdigt, når alle Post It's er i en gruppe og ingen ønsker at rykke rundt på grupperingerne - en gruppe kan godt bestå af én Post It.

15.

Opstilling af Attributter - Step 4

4. Navngiv alle grupper.

- Undlad helst hele sætninger
- Det er lovligt at splitte og sammensætte nye grupper i dette trin af processen
- En gruppe kan have mere end ét navn, så forsøg at give alle grupperinger af Post It's et navn
 - Undlad kun at give en gruppe et navn, hvis en anden deltager allerede har skrevet den eksakte formulering du tænkte på
 - ... Diskussion og snak er stadig ikke tilladt :-)

16.

Opstilling af Attributter - Step 5

5. Stem på hvilke af de nyformede grupper, inden for hver overkategori, som er mest vigtig og som bedst besvarer fokusspørgsmålet fra tidligere

- Noter individuelt navnene på de tre grupper man personligt finder vigtigst. Har en gruppe mere end ét navn, noter da det navn du finder bedst beskrivende

Fokus spørgsmål

Hvilke attributter og karakteristika beskriver en diabetiker og gør sig gældende under en hvilken som helst evaluering?

17.

Opstilling af Attributter - Step 5

5. Stem på hvilke af de nyformede grupper, inden for hver overkategori, som er mest vigtig og som bedst besvarer fokusspørgsmålet fra tidligere

- Noter individuelt navnene på de tre grupper man personligt finder vigtigst. Har en gruppe mere end ét navn, noter da det navn du finder bedst beskrivende
- Når alle har noteret navnene på grupperne rangere man dem fra mest vigtig til mindst vigtig.

Fokus spørgsmål

Hvilke attributter og karakteristika beskriver en diabetiker og gør sig gældende under en hvilken som helst evaluering?

18.

Opstilling af Attributter - Step 5

5. Stem på hvilke af de nyformede grupper, inden for hver overkategori, som er mest vigtig og som bedst besvarer fokusspørgsmålet fra tidligere

- Noter individuelt navnene på de tre grupper man personligt finder vigtigst. Har en gruppe mere end ét navn, noter da det navn du finder bedst beskrivende
- Når alle har noteret navnene på grupperne rangere man dem fra mest vigtig til mindst vigtig.
- Når alle har rangeret finder man grupperne på tavlen og sætter kryds efter vigtighed:
 - Vigtigst: 3 kryds
 - Næstvigtigst: 2 kryds
 - ... "mindst vigtig": 1 kryds

Fokus spørgsmål

Hvilke attributter og karakteristika beskriver en diabetiker og gør sig gældende under en hvilken som helst evaluering?

19.

Opstilling af Attributter - Step 6

6. Rangering af de 'mest' vigtige grupper

Alle grupper med stemmer flyttes op. Gruppen med flest stemmer øverst og derefter næsthøjest antal stemmer osv.

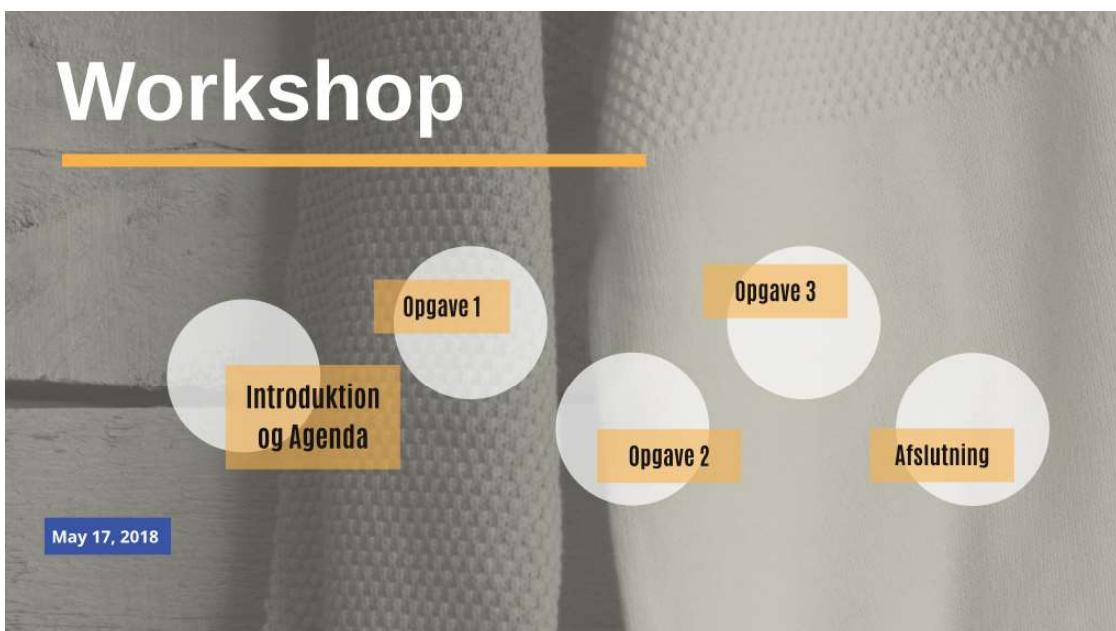
Læs grupperne og deres indhold. I må gerne diskutere tingene nu.

- Identiske kategorier, kan slås sammen.
 - Grupperne skal være så identiske at de kan erstatte hinanden.
 - Grupperne kan kun slås sammen, hvis der er enstemmighed omkring det.

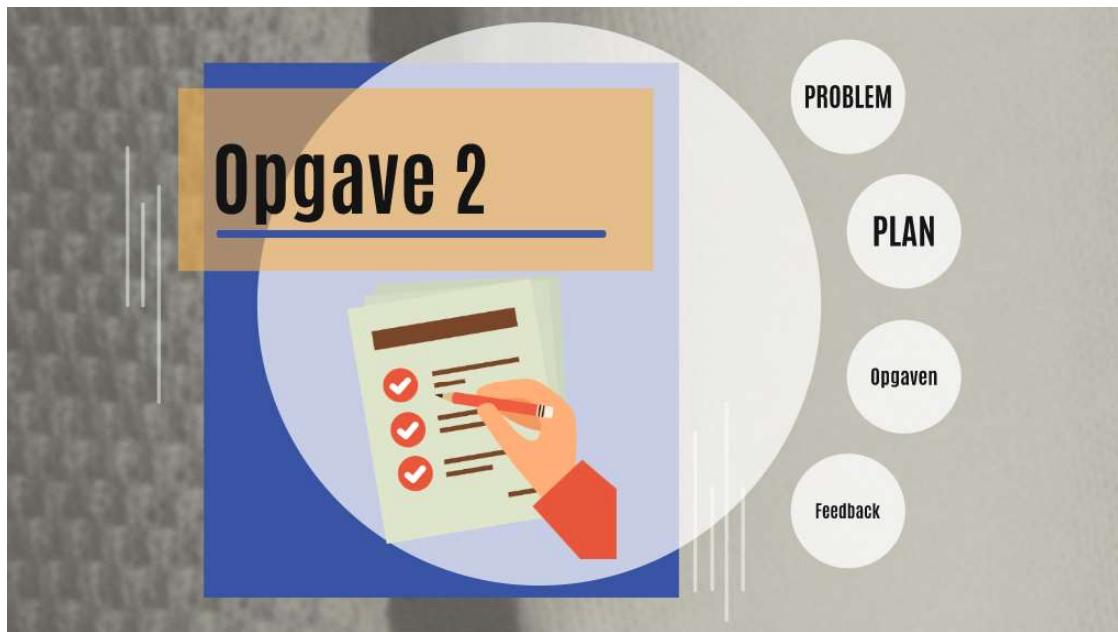
20.



21.



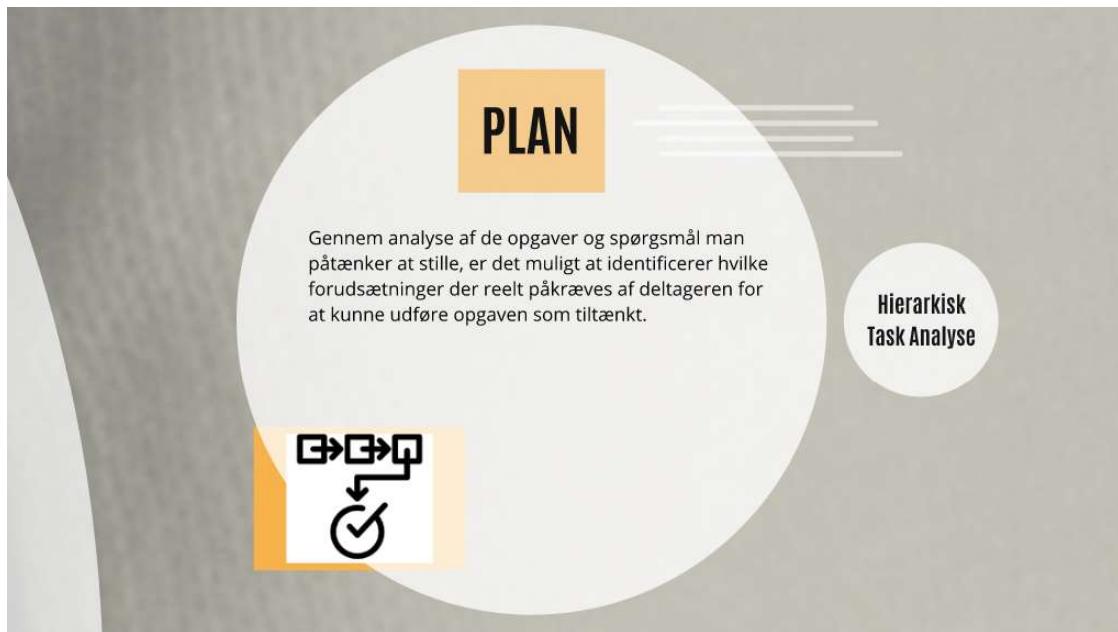
22.



23.



24.



25.



26.

Familiarisering med Task Analyse

Lav individuelt en task analyse af opgaven:

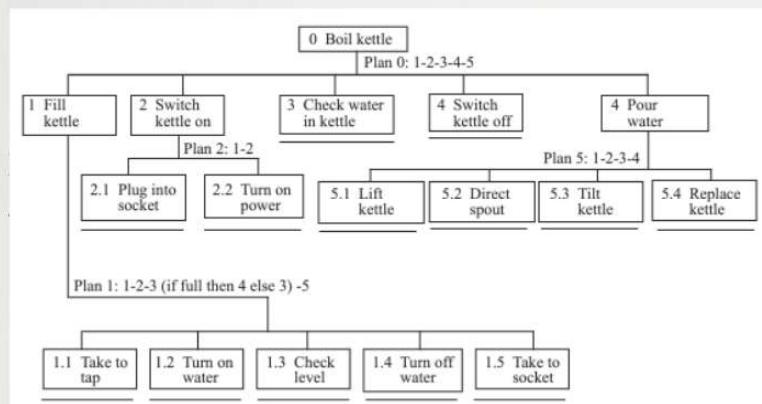
- Lav en kop te

1. Definer det **overordnede formål** for opgaven (I dette tilfælde: At lave en kop te)
2. Definer alle **undermål** for opgaven
3. Nedbryd alle undermål indtil et/flere **operationelle krav** til opgaven er defineret
4. Definer alle **påkrævede/indflydende attributter og karakteristika** for hvert af de operationelle krav der er blevet fundet frem til.
5. Definer den påkrævede rækkefølge for operationerne. Husk at nummerere alle post It's. Fx. det første overordnede formål som 1, undermålene som 1.X og de operationelle krav 1.X.X

27.

Familiarisering med Task Analyse

Lav individuelt en task analyse af opgaven:



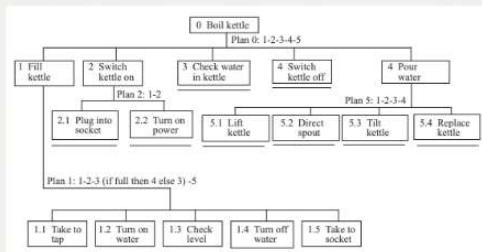
28.

Hierarkisk Task Analyse

Sammen i gruppen skal i lave en Hierarkisk Task Analyse

- Tag udgangspunkt i FE048.
- Følg nedenstående punkter. Detaljegraden for outputtet af analysen skal være i ca. samme detaljegrad som eksemplet på figuren.

1. Definer det **overordnede formål** af hver opgave (f.x. Injicer 20 units)
2. Definer alle **undermål** for hvert enkelt opgave
3. Nedbryd alle undermål til mindre undermål indtil et/fleste **operationelle krav** til opgaven er defineret
4. Definer alle **påkrævede/indflydende attributter og karakteristika** for hvert af de operationelle krav fra den hierarkiske task analyse.
5. Definer den påkrævede rækkefølge for operationerne. Husk at nummerere alle post It's. Fx. det første overordnede formål som 1., undermålene som 1.X og de operationelle krav 1.X.X

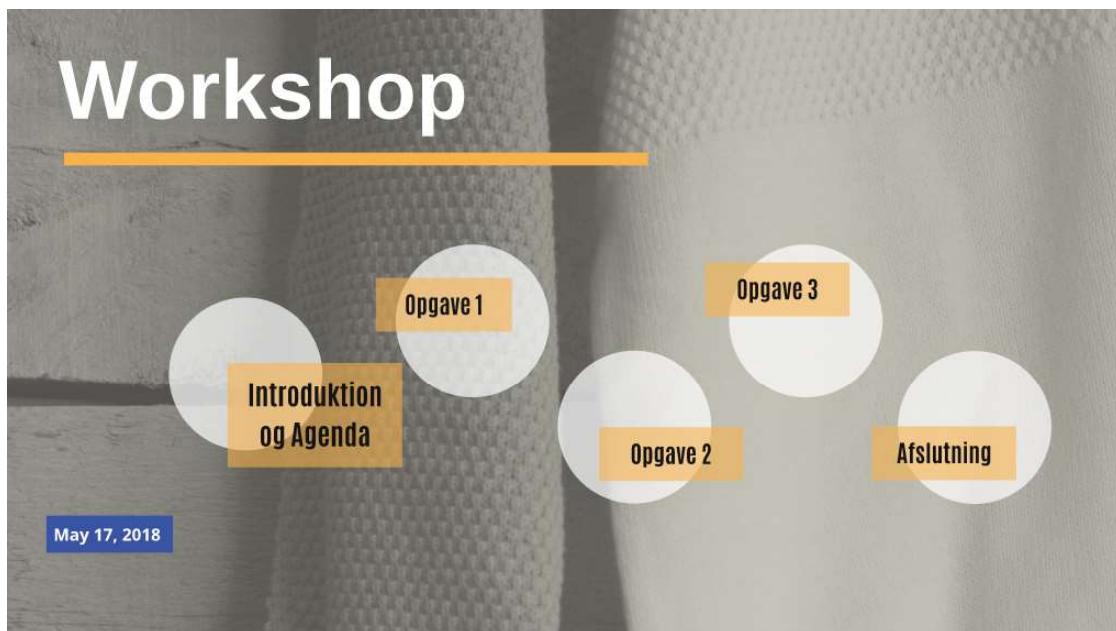


29.

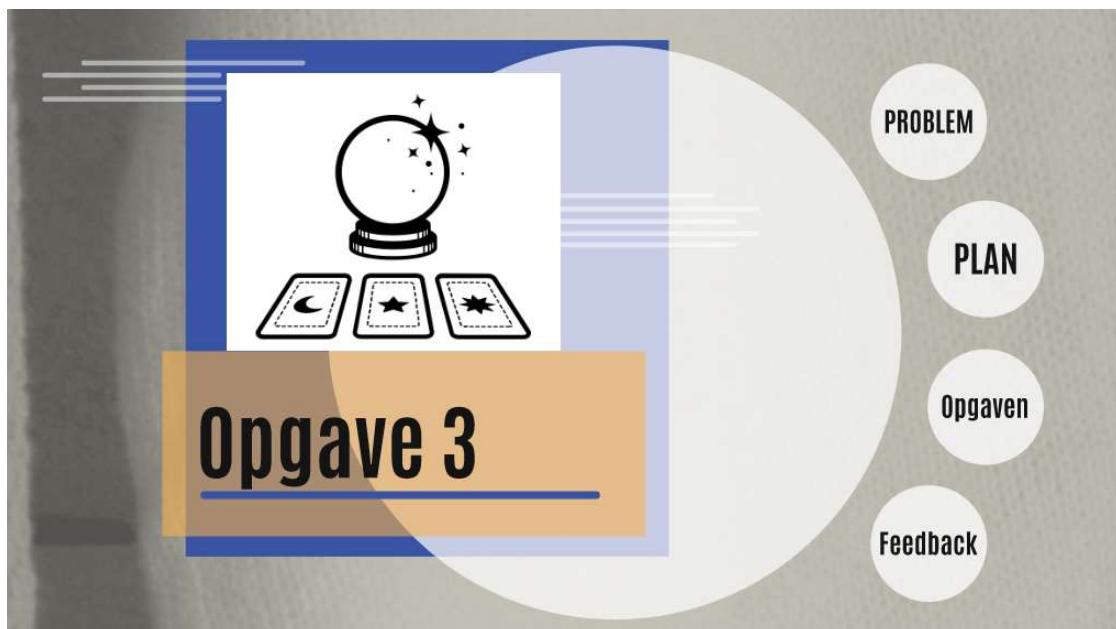
Spørgsmål og feedback



30.



31.



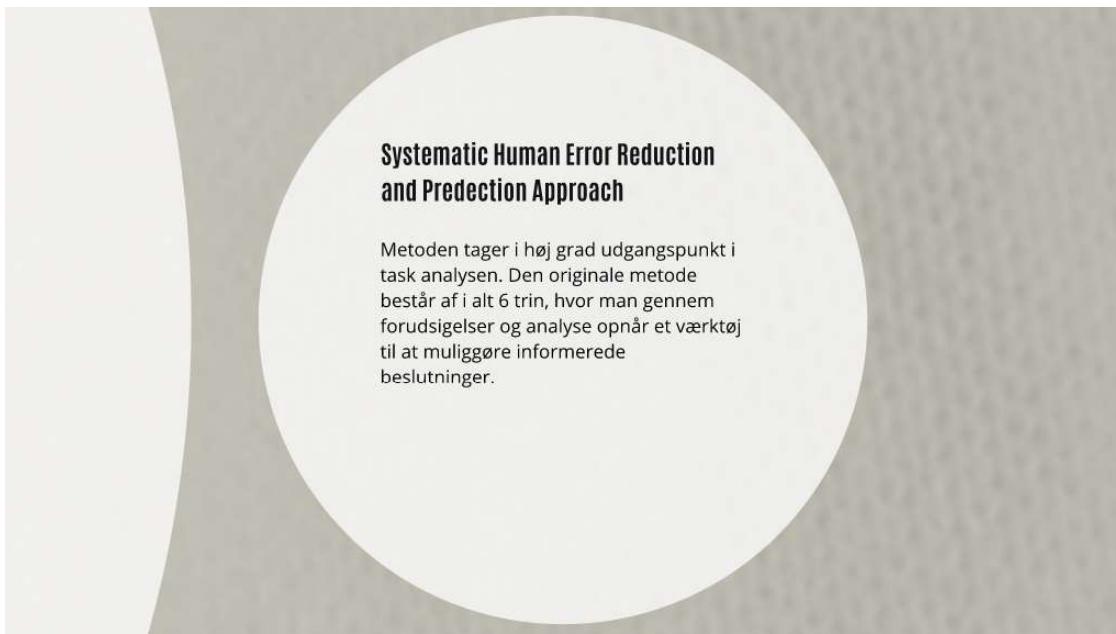
32.



33.



34.



35.



36.

SHERPA - Step 1

Task klasficerings

Som start skal alle de operationelle krav fra task analysen klasficeres. Dette gøres gennem en veletableret taxonomi. Taxonomien består af 5 kategorier, som er:

- Action (e.g., pressing a button, pulling a switch, opening a door)
- Retrieval (e.g., getting information from a screen or manual)
- Checking (e.g., conducting a procedural check)
- Selection (e.g., choosing one alternative over another)
- Information Communication (e.g., talking to another party)

Klasficer alle de operationelle krav i analysen der blev enighed om at fortsætte med i sidste opgave.

37.

SHERPA - Step 2

Fejl identification

Hvert operationelt krav bør være klasficeret. Alt efter klasficerings, skal disse nu tildeles endnu en taxonomi.

Tilfør fejl taxonomien til alle operationelle krav. Hvis ingen af de prædefinerede taxonomier passer, foreslå da selv en.

Tilføj endeligt en kort beskrivelse for fejl; hvordan og hvorfor?

Action errors	Checking errors
A1 Operation too long/shprt	C1 Check omitted
A2 Operation mistimed	C2 Check incomplete
A3 Operation in wrong direction	C3 Right check on wrong object
A4 Operation too little/much	C4 Wrong check on right object
A5 Misalign	C5 Check mistimed
A6 Right operation on wrong object	C6 Wrong check on wrong object
A7 Wrong operation on right object	
A8 Operation omitted	
A9 Operation incomplete	
A10 Wrong operation on wrong object	

Retrieval errors	Communication errors
R1 Information not obtained	I1 Information not communicated
R2 Wrong information obtained	I2 Wrong information communicated
R3 Information retrieval incomplete	I3 Information communication

38.

SHERPA - Step 3

Konsekvensanalyse

Hver forudset fejl har en eller anden konsekvens.

Beskriv konsekvens(erne) for de fejl der er blevet forudset i det tidligere step.



39.

SHERPA - Step 4

Genopretningsanalyse

Når man kender potentielle fejl og konsekvenserne derved, kan man forudsige hvordan disse kan genoprettes eller helt undgåes.

Det kan bl.a. indebære småændringer i design af prototype eller opgave. En fejl kan nogle steder gøre det umuligt at fortsætte, hvoraf potentialet for genoprettelse ikke eksistere.

1. Afgør om det er muligt at genoprette de forudsete fejl.
2. For de fejl der kan genoprettes/undgåes skrives en 'genoprettelses-strategi'

40.

SHERPA - Step 5

Sandsynlighedsanalyse

Alle forudsete fejl skal tildeles en sandsynlighed for at ske. Det gøres ved at bedømme dem som lav (L), mellem (M) eller høj (H). Måden dette gøres på sker efter følgende regler:

1. Hvis fejlen ikke er sket i tidligere situationer, bedømmes sandsynligheden som lav
2. Hvis fejlen er sket før i tidligere situationer bedømmes sandsynligheden som mellem
3. Hvis fejlen ofte er sket i tidligere situationer bedømmes sandsynligheden som høj

41.

SHERPA - Step 6

Betydningsbetydnings

Med henblik på potentielle fejl og konsekvenserne deraf, er det muligt at vurderer hvor kritisk fejlen vil være. Vil fejlen kunne stoppe deltageren i at gennemføre, vil det introducerer en bias eller vil det umuliggøre at opnå den data man ønsker.

Hvor kritisk fejlen er bedømmes med samme skala som konsekvenserne (L, M, H). En kort argumentation tilføjes efterfølgende.

42.

SHERPA - Step 7

Sample konsekvensanalyse

Med udgangspunkt i de attributter og karakteristika der er blevet vurderet nødvendige eller indflydende på den enkelte opgave, er det muligt at opstille konsekvenser. Det vil kunne afgøre hvilken betydning det vil have at bruge en sample der ikke opfylder disse.

Prøv at samle jer og diskuter konsekvenserne af at bruge en anden sample end diabetikere for FE048.

43.

SHERPA - Step 8

Betydningsbedømmelse

Konsekvenserne ved at bruge en alternativ sample muliggør at bedømme hvor kritisk det er at deltagerne i samplen opfylder de behov der initierende er påkrævet. Det afhænger øjensynligt af konsekvensen.

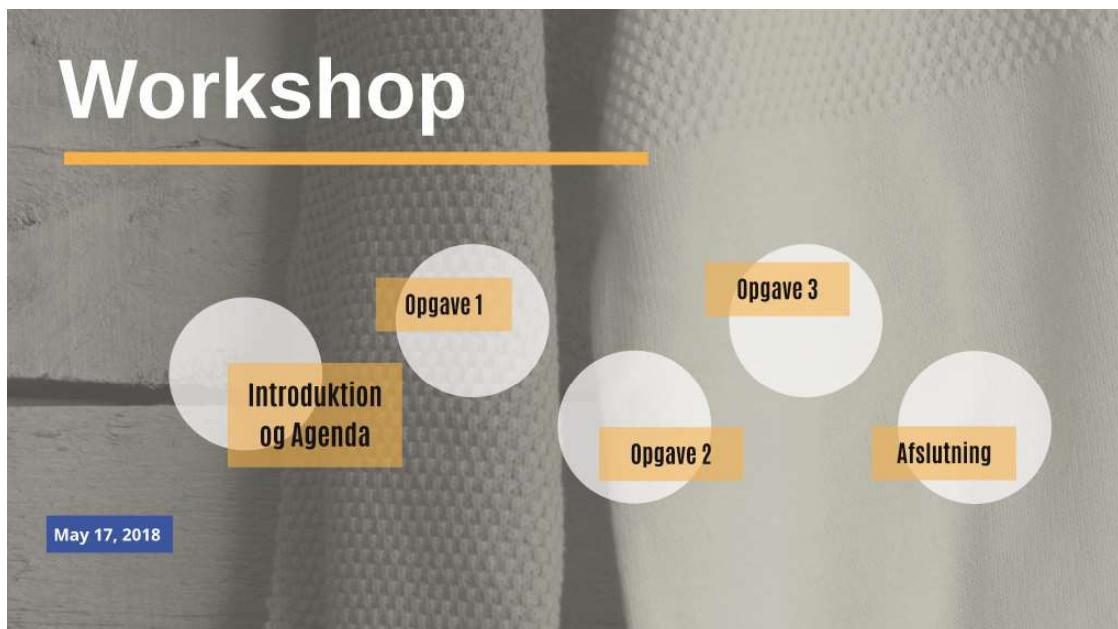
Betydningen bør holdes op mod objectives af evalueringen samt den enkelte opgave. Det bør overvejes om brugen af en anden sample kunne stoppe deltageren i at gennemføre, vil introducerer en bias eller vil det umuliggøre at opnå den data man ønsker.

Samme skala, som tidligere benyttes (L, M, H). En kort beskrivelse for hvorfor betydningen bedømmes som den går, gives derefter.

44.



45.



46.



47.



48.



49.



50.

Presenters Notes

Slide 1	<p>Velkommen - dejligt at se jer alle sammen igen.</p> <p>Jeg vil ikke sige så meget andet end at vi gennem de næste tre timer vil skulle bruge tre forskellige metoder.</p> <p>Der er planlagt pauser undervejs, men det finder I ud af.</p> <p>Jeg regner med at vi slutter omkring kl. 16. Jeg ved nogle af jer skal gå lidt før og en enkelt kommer også senere, men det gør ikke noget.</p> <p>Men lad os komme i gang.</p>
Slide 2	<p>Teori og litteratur siger ikke så meget, så vi skal i dag prøve at bruge en proces der måske kan afhjælpe problemerne der er omkring sampling.</p> <p>I specialet har jeg undersøgt hvad der gøres i andre virksomheder, samt fået inspiration til hvordan man kan gøre emnet an.</p>
Slide 3	<p>Gennemgå agenda - hvis der er spørgsmål undervejs, så stil dem endeligt. Det er ikke sikkert jeg vil svare på det med det samme, men der er rig mulighed for spørgsmål og lignede i de feedback-sessions der er planlagt efter hver opgave.</p>
Slide 4	<p>Præsenter overordnet de 3 opgaver, som skal løses henover de 3 timer.</p> <p>Analyserne og værktøjerne der bliver brugt er udvalgt på baggrund af teori.</p> <p>Forhåbentlig vil processen kunne være afgørende for ens sample og test.</p> <p>Endeligt er meningen med i dag, at i benytter metoderne og at vi efterfølgende diskuterer dem.</p>
Slide 5	<p>Præsenter brugsområdet for processen.</p> <p>Nævn at der typisk ved sampling ikke er nogle 'regler' men at man typisk bare sigter efter at få så repræsentativ en sample som muligt.</p>
Slide 7	<p>Den første opgave omhandler den population man udvikler til.</p>
Slide 8	<p>Præsenter problem - det er muligvis indlysende at en population er unik, men hvordan omgår man så det i en evaluering?</p>
Slide 9	<p>Plan: Fastslå de attributter og karakteristika der reelt udgør en specifik population; med andre ord, de ting der gør population unik.</p>
Slide 10	<p>User profiling og personaer er to metoder til at opnå en indsigt i de brugere man designet til.</p> <p>En gennemarbejdet persona eller user profile er ikke opnåligt henover 3 timer, men vi vil forsøge at lave noget af grundarbejdet.</p>
Slide 11	<p>Vi kommer til at tage udgangspunkt i udviklingen af personaer. Udviklingen af en rigtig god persona tager virkelig lang tid, så vi kommer ikke til at arbejde med den fulde proces. Hvis ressourcerne og tiden er til det, kan man sagtens drage nytte af en god persona. Den kan endda bruges i andre sammenhænge, end kun sampling.</p>
Slide 12-13	<p>Brainstorm attributter indenfor de 5 typer af adfærds-variabler.</p> <p>(Gule Post It's i denne fase)</p>

10 minutter i alt, prøv at få skrevet mindst én til hver af kategorierne.

Slide 14	Læs slide Stadig gule post it's, hvis der er tilføjelser
Slide 15	Læs slide Pointer at der endnu ikke gives navne til grupperingerne. "Take two items that seem like they belong together and place them in an empty portion of the wall, at least 2 feet away from any other sticky notes. Then keep moving other like items into that group." "Feel free to move items into groups other people create. If, when reviewing someone else's group, it doesn't quite make sense to you, please feel free to rearrange the items until the grouping makes sense." "You're to complete this step without any discussion of the sticky notes or the groups. Every item has to be in a group, though there are likely to be a few groups with only one item."
Slide 16	Jeg vil nu at I navngiver hver gruppe. Læs indholdet af alle grupper og find det navn, som man bedst synes repræsentere gruppen. Dette gøres på Blå Post It's. Blå post it's "I want you to now give each group a name. Read through each group and write down a name that best represents each group on the new set of sticky notes I just gave you." "A name is a noun cluster, such as 'Printer Support Problems'. Please refrain from writing entire sentences." "As you read through each group, you may realize that the group really has two themes. Feel free to split those groups up, as appropriate." "You may also notice that two groups really share the same theme. In that case, you can feel free to combine the two groups into one." "Please give every group a name. A group can have more than one name. The only time you're excused from giving a group a name is if someone has already used the exact words you had intended to use."
Slide 17-19	Noter på sedlerne, navnene på de tre mest vigtige grupper i hver kategori. Man må ikke vise navnene til de andre deltagere. Efter navnene er færdige: Lav nu en prioritering af de grupper i har valgt. Efter prioritering: Op og sæt krydser (6 krydser for hver overkategori (3 til den vigtigste, 2 til den næstvigtigste, 1 til den mindst vigtige))

Lad os starte med:

Aktiviteter
Attituder
Aptituder
Motivation
Færdigheder

Slide 20 Flyt alle grupper med stemmer op øverst - højeste antal stemmer og ned.

"We now need to see if there are any groups that we should combine. You can nominate two groups that you think are the same thing."

"We'll then take a preliminary vote, to see if anyone thinks they aren't the same. If anyone believes they are different, we'll spend a little time discussing why they believe that."

"After the brief discussion, we'll take a final vote. That vote needs to be unanimous for us to combine the items and their scores."

"Remember, the two groups being considered need to be identical. That means you could substitute one for the other. A group that's a subset of the other group does not qualify for combining."

Slide 21 Tag spørgsmål først

Feedback.

Hvad synes I om det at fastlægge attributter og karakteristika?

Tilfører det noget ny viden?

Finder I det brugbart?

Har I forslag til ændringer, tilføjelser eller ting der er overflødige?

PAUSE (5-10 min alt efter tidsplan)

Hjælp til at samle data ind (Det skal være muligt at genbesøge)

DISKUSSION: Motivation: Er det muligt at skabe samme motivation i et evalueringsscenarie som brugeren oplever til hverdag??????

Slide 23 Opgave to omhandler evalueringens rolle og betydning for hvilken sample man kan/vil benytte.

Slide 24 Typen af evaluering fylder også meget. Det betyder samtidig at forståelse for brugerens ikke er nok, da problemet skal observeres fra flere forskellige synspunkter.

Grunden til at typen af evaluering har betydning skyldes hovedsageligt forskellige samples har forskellige forudsætninger for at svarer.

Slide 25 Gennem analyse skal det forsøges at skabe en sammenhæng mellem evaluering og sample. Lige før fandt vi de attributter og karakteristika der beskriver gruppen. nu vender vi det hele på hovedet og kigger reelt set på, hvad evalueringen påkræver af brugeren.

Det er også muligt at en opgave bliver påvirket af brugerens viden, motivation eller lign. så det er ikke kun brugerens færdigheder vi kigger på.

Slide 26 For at gøre dette, bruger vi den hierarkiske task analyse.

Læs slide.

Formålet er at kigge på hvad det reelt er vi beder en deltager om at gøre, når vi stiller opgaven. Det kan virke meget simpelt, men typisk kræver opgaverne mere af brugeren, end først forventet.

Slide 27-28 Jeg vil gerne hvis i individuelt vil lave en task analyse på opgaven: "Lav en kop te."

Der er ingen forventninger til det endelige resultat, men det foreslås at følge punkterne. Er der spørgsmål, undervejs, så sig endelig til.

Efter familiarisering:

Få deltagerne til at se hvor forskellige/ens deres analyser er.

Vis hvor meget det egentlig kræver bare at koge en kedel med vand

Slide 29 ***Uddel print af FE048***
*** Uddel A3 og Post It's***

Læs slide og giv eksemplet. Meget vigtigt at give eksempel for hvert af punkterne på slide.

Præsenter opgaven

Farven på teksten angiver hvilken farve post it's det skal noteres på.

Overordnede mål: Lyserøde Post It's

Undermål: Blå Post It's

Operationelle krav: Grønne Post It's

Attributter/Karakteristika: Gule Post' It's

Slide 30 Tag spørgsmål først

Feedback.

Hvad synes I om det?

Tilfører det noget ny viden?

Finder I det brugbart?

Har I forslag til ændringer, tilføjelser eller ting der er overflødige?

PAUSE (5-10 min alt efter tidsplan)

	Hjælp til at samle data ind (Det skal være muligt at genbesøge)
Slide 32	Den sidste opgaver for i dag, indebærer at bruge de data vi har opnået tidligere. Primært omhandler det forudsigelser. Forudsigelserne tager udgangspunkt i jeres ekspertise inden for området.
Slide 33	Vi kender populationen og evalueringen, men der der stadig ikke et svar på hvad hvilke mulige fejl der kan ske, konsekvenserne deraf samt hvor kritisk det egentlig er. Gennem det sidste skridt skal vi prøve at lægge en strategi der kan bruges fremadrettet.
Slide 34	Omsætte resultater til et værktøj og argumentationsmiddel for både produkt-design, evaluering og sample
Slide 35	Gennem analyse og forudsigelser vil man forhåbentlig kunne tage en masse informerede beslutninger. Værktøjet kan i høj grad bruges til at: Identificerer fejlkilder i forbindelse med opgaver og sample Reducerer disse Undgå dem.
Slide 36	Metoden her og de forudsigelser I vil komme til at lave, baseres på den tidligere analyse af FE048.
Slide 37	Første step til at forudsige fejl og konsekvenser er at klassificerer de operationelle krav. Læs slide: På en Lyserød Post It noteres nummeret på den task man klassificerer (f.x. 1.X.X) så det er muligt at backtracke. Nummeret skal gerne fremgå på opgaven i Task Analysen fra før. Endeligt vil jeg meget gerne at i på alle post it's i notere hvilket step den tilhører - udelukkende for at jeg kan holde styr på dem efter workshoppen :-)
Slide 38	Næste skridt er at identificerer fejltypen. På en blå Post It skrives en af fejltyperne. Fejltypen angives både med forkortelsen og den korte beskrivelse dertil, eksempelvis "A1 - Operation too long/short." Endeligt skrives en kort beskrivelse af hvordan og hvorfor fejlen potentiel kan ske. Det behøver ikke være mere end hvad der kan stå på en post it.
Slide 39	De fejltyper i har noteret skal nu tildeles en konsekvens. Hvad sker der reelt hvis fejlen opstår? Et operationelt krav kan have én eller flere konsekvenser. Disse skal beskrives kort. Beskriv konsekvenserne på Grønne post it's.

Slide 40 Læs slide

Genoprettelses-strategien behøver ikke være mere end ti ord. Det kan eksempelvis være 'Instead of asking participant to dial, set it to a predefined amount', hvis nu man forudser at dial vil være problematisk.

Løsningen behøver ikke være optimal, men bør noteres. Hvis løsningen er en naturlig constraint: fx. hvis man ikke får hætten af sin pen, så vil denne fejl forhåbentlig rettes op ved et af de senere trin - det er eksempelvis ikke muligt at injicere uden hætten er taget af, hvoraf dette tvinger deltageren til at gå tilbage til denne fejl.

Genoprettelses-strategien skrives på en gul post it.

Slide 41 Læs slide

På den blå Post It' som beskriver fejlen skrives bedømmelsen.

Slide 42 Læs slide

På den grønne Post It', hvor konsekvensen af fejlen er skrevet, skrives bedømmelsen af hvor kritisk det vil være.

Argumentationen kan tilføjes på en post it nedenunder konsekvensen.

Slide 43 Med udgangspunkt i den viden der er blevet samlet gennem opgaverne skal vi forsøge at forudsige hvilke konsekvenser det vil have, hvis man ikke benytter den reelle sample.

Det indebærer viden fra opgaven omkring populationen i sig selv. Det indebærer at bruge viden fra Task Analysen - især hvilke forudsætninger der påkræves i den enkelte opgave. Og endeligt kræver det viden fra denne sidste opgave.

Prøv at samle jer og diskuter konsekvenserne af at bruge en anden sample end diabetikere.

Noter endeligt disse konsekvenser på lyserøde post it's. Husk venligst også at skrive Step 7

Slide 44 Det sidste skridt i analysen er at bedømme betydningen af at bruge en anden sample end den man designer til.

Det skal gøres for hvert enkelt opgave, da antallet af steder hvor det vil være muligt at bruge en anden sample klart vil have indflydelse på det endelige valg.

På en ny blå Post It' bedømmes betydningen og samtidig given en kort begrundelse for hvorfor bedømmelsen er sådan.

Husk venligst at notere step 8 på denne.

Slide 45 Tag spørgsmål først

Feedback.

Hvad synes I om det at fastlægge attributter og karakteristika?

	Tilfører det noget ny viden?
	Finder I det brugbart?
	Har I forslag til ændringer, tilføjelser eller ting der er overflødige?
	PAUSE (5-10 min alt efter tidsplan)
	Hjælp til at samle data ind (Det skal være muligt at genbesøge)
Slide 47	N/A
Slide 48	Læs slide. Min oprindelige tanke var selv at gennemgå den her proces, som i nu har taget. Ud fra det ville jeg have identificeret en anden sample og have forsøgt at påvise at resultaterne muligvis kan opnås uden en repræsentativ sample. Præsenter begrebet 'Burden of proof'. Selvom metoderne er anerkendte, er det ikke ensbetydende med at højere autoriteter (FDA) vil acceptere det.
Slide 49	De oprindelige værdier af metoderne bibeholdes: læs slide.
Slide 50	Tag spørgsmål først Feedback. Tilfører det noget ny viden? Finder I det brugbart? Er tankegangen med de tre metoder forståelig? Har I forslag til ændringer, tilføjelser eller ting der er overflødige? I wish & i Liked

Case for Task Analysis

Exercise 1

- **Attach**

Please attach the device to the pen.

*[Observe how the participant intuitively orientates the device.
Any difficulties?]*

- **Ease?**

How was that? Easy/difficult?

- **Accidentally Come Off?**

What would make this device accidentally come off the pen?

If so, tell me about these situations. How likely are they?

- **Orientation**

I noticed that you held/orientated/tried to..... Why was that?

How did you know that the device attaches in this way?

- **IFU as last course of action**

[If they are not able to attach the device, hand them the IFU extracts]

Exercise 2

- **Prepare for injection simulation**

You will be simulating an injection in a moment. Obviously, I do not want you to inject into yourself. Therefore, I would like you to attach this injection pad to your body – where you would normally inject yourself.

These are the needles we are going to use
[Place a cup of needles on the table].

- **Simulation Scenario**

I would like you to imagine that you have downloaded the app which is recommended on the box and you have set it up with the device so the Device & App are communicating.

[Open App, and put it on the table in front of the participant]

- **Injection**

Now, it is time for your injection. Please take a dose of 20 units.

How was that?

- [Tech box command: 20 units] Accurate
- [Observe how they hold the Device. One or two hands? Do they cover the dose counter?]

Exercise 3

- **2nd Injection: Different Place**

Let's imagine that you need to do another injection. Do you sometimes inject other places than [injection location in exercise 2]? Okay, could you please attach the cushion there? (If no other location, ask them to do the injection a different place than the first one)

Can you please do another injection of 1 unit.

Tech box command: 2 units Inaccurate

[Observe how they hold the Device]

How was that? Different than the first time?

[Observe if they are looking at the App, and if they are noticing that it did not transfer the correct dose]

You mentioned... I noticed...

[Explore on inaccuracy reaction]

- **Reaction to Inaccuracy?**

Imagine this kind of error will continue to occur, in other words sometimes the reading will be one unit more or one unit less than was actually administrated.

- **Injection Confidence/Comfort**

How comfortable/confident do you feel performing an injection with this device? How does it compare to your normal injector?

I noticed during the first exercise you hold the device....

[Ask follow-up questions related to how they hold the Device in Exercise 1 and 2, if not already mentioned]

Exercise 4

Empty Pen

Let's imagine that you have used this device and pen for a while. There is no more insulin in the pen. Go ahead and remove the accessory device from this pen and attach it to this new pen.

[Observe how the participant intuitively orientates the device, when doing it again. Is anything different from the first time?]

How was that? How was it to detach the device? What about attaching it again?

End At the end of this interview, . . .	Very Unsure	Somewhat Unsure	Somewhat Sure	Very Sure
I know how to use this device.				
I know how to attach & detach this device as intended.				
I would want to use this device regularly.				

Figure 1 "Confidence in Device" Questions in Worksheet at Start of Interview

Worksheet: End of Interview

Let's take a look at the worksheet you filled out at the beginning of our interview.

Answers?

How would you answer these questions now that you've spent some time interacting with the device?

Changes?

What makes you change your answers?

Results from Assignment 1

Table H.1: Results from the first assignment of the workshop.

Activities		
Categories	Score	Content
Lifestyle	0	Unhealthy lifestyle Hard to administer the disease as a teenager They might do nothing at all
Need for care	2	Typically has a support who is indispensable in order to feel safe
Control	4	Has everything necessary for treating diabetes under control Monitors food intake Makes calculations and systems to keep track of the disease They balance many parameters (food, exercise, medicine) They keep track on different kinds of insulin
Medical treatment	6	Takes injection at meals Takes injections before going to bed Inject insulin multiple times a day They take pills every day They measure blood sugar multiple times a day They control a pump They often carry insulin pens with them
Attitudes		
Categories	Score	Content
Fear of needles	0	Fear of needles Troublesome and has to be done multiple times a day General opponent/fear of injecting
Expectations to the future	0	The development is too slow Wishes to be cured
Contact with experts	0	Too little contact and understanding of doctors and specialists
Positive approach	1	Happy for pen-shaped device Grateful (medicine being a life-saver) Happy to have the ability to inject
Ease of burden	0	Won't trouble others with their disease Annoying reminders of the disease
Social burden	0	Feels lonely with the disease Is perceived as dumb and obese by doctors Health care systems are flawed
Necessary treatment	0	They know the medicine is vital They know to exercise

Continued on next page

Continued from previous page

Inconvenience	1	They know to measure blood sugar They find it a hassle to start on medication and they wish to stop taking it
---------------	---	--

Aptitudes

Categories	Score	Content
Lower education	0	Generally lower educational level
Resignedly	0	Possibly lower understanding of health than average
Society based training	0	No training Self-taught Taught by a health care physician Is offered education and courses regarding diabetes

Motivation

Categories	Score	Content
Identity	3	They want to drop the medication (type 2) They want to forget about the illness Interested in technology, as they want to feel 'normal'
Necessity	0	Necessity for surviving and health Treatment is necessary. They can feel the need for care
Reasonable sense	0	The physician tells me i have to
Anxiety	3	They are motivated, as they don't want to get worse Anxiety for hyphos Afraid of seizures
Related disease	0	Has other disease due to diabetes

Skills

Categories	Score	Content
Super Users	2	They usually inject insulin in many years and become 'Super users' Injections experts Daily training with devices
Experience	0	Has seen family members take injections before

Chapter I

The KJ-Technique

The KJ-technique is designed for establishing priorities in groups. It is especially useful when limited by resources and it tends to be quite accurate. The strength of the method is to use a lot of subjectively and opinionated data as input and processing it to an objective outcome. It is comparable to affinity diagramming and the previous described method, Yin's Five Phased cycle and it shares some of the aspects regarding the analysis of qualitative data in general (Spool, 2004).

As for many other methods, the process of using the KJ-technique is divided into steps. There's a total of eight (8) steps in the KJ-technique, but only six (6) of them are to be conducted by the group. The first two (2) steps is all about establishing the framework for the workshop and to organize the group of participants that are intended to participate. Below all steps are described and they all follow the procedure described by Spool (2004).

Procedure of using the KJ-Technique

Determining a Focus Question

The initial step of the KJ-technique is to determine a focus question. The focus question is what is trying to be answered throughout the whole sessions. It is the focal point of the 'workshop' and is the main thing participants are going to answer.

There really is no right and wrong in this step, but the most important questions should always be prioritized. It is important to get it right, as it will have an impact, when the participants have to start working with it.

Organizing the Group

The use of KJ usually takes around an hour, which people have to be informed about. Spool (2004) states that a diversity of participants is to prefer, as this enables a variety of perspectives. One rule, however, should be that only people with connection and association to the subject being treated should be invited. There's no reason having confused participants who won't be able to provide their perspective, if they in reality doesn't have any on the subject being treated.

The Start of the Workshop - Putting Opinions on Post It's

All participants are simply asked to brainstorm and note down as many opinions or facts onto sticky notes. The only requirement is that they have some kind of association to the focus question and that it is done individually. Having the participants note down individually allows everybody to provide input. When not doing it individually but in a group instead, there's a chance of people taking the lead. If this is the case, some people might be suppressed. In this way, the leader of the group dictates what is 'good enough' to be noted down.

Putting Post It's on the Wall

The participants are asked to place all notes from the brainstorm onto the wall (or any other usable surface). This is done in random order and there's no system - yet. The participants are asked to read what others have noted. If they get inspired and think up new aspects, they are allowed to create new notes to be put up among the other items.

Grouping Items

When everybody have noted down every single aspects they could come up with, the participants are asked to group the items. The rules for grouping two items is that they should have similar content. The step of grouping is still done individually, and if someone disagrees with a grouping, they can simply split it and do the group they think of. This really calls for discussion, but this isn't allowed. Instead, the participants have to obtain unanimous consent. The step of grouping isn't finished until everybody agrees.

Naming the Groups

Next, the participants are asked to name each of the groups they just made. Each participant is asked to name each group. The only reason for not naming a group is, if another participant has named the group exactly as one would have done.

In this step, it is also allowed to split groups, if it is realized that the content of the group doesn't appear as similar as first. Additionally, it is also allowed to combine groups, if a participant realizes that the content of two or more groups are similar.

Through this step it still isn't allowed to discuss nor talk with the other participants. While the goal of this step is to name the groups, there's also a hidden agenda: the final review. Forcing all participants to read the content of every group is also a review of everything, and it makes the participants consider every aspect on the wall.

Voting for the Most Important Groups.

Second to last, the participants have to vote for the groups they find the most important. The way to judge this, is to subjectively choose the three most important aspects related to the focus question, that was presented in the beginning; what aspects answers to focus question the best.

The process of voting is split into three stages. The first is to make the participants note down the three most important groups on a piece of paper. If a group has more than one name, the name that one finds best fitting, is the one to note down.

When all participants have noted down three groups, they are asked to prioritize them. When everybody has done this, the votes are recorded by asking the participants to set three x's on the first priority, two for the second and one for the last priority. This is done on the sticky note on the wall, in order for everybody to see the votes. This is still done without any discussion or talk amongst the participants.

Raking the Most Important Groups

Having all the groups with votes, these are now ordered. The groups with the highest amount of votes at the top. The participants are asked to gather around, while the groups are read out loud. Some groups might represent identical priorities, and the participants are allowed to consider combining these groups. The only way for two groups to be combined is if the group unanimously agrees. Furthermore, it is mentioned that a grouping at this stages means that the two are **identical**. It's not a subset - a rule of thumb is that the two groups in questions should be able to substitute each other completely.

At this point discussion is allowed. The viewpoints for and against combining groups can be spoken. If two groups are to be combined, the scores from each individual group are combined and moved higher up the list. When everybody agrees that there's no more to be changed, the exercise is finished. The outcome should be a prioritized list of aspects answering the focus question.