TR
Technical Report

CACHET Unified Methodology for Assessment of Clinical Feasibility

Establishing clinical evidence for the feasibility of personal health technology during design, development, and pilot testing

Jakob E. Bardram

cachet | Copenhagen Center for Health Technology

CACHET
Copenhagen Center for Health Technology

Technical University of Denmark
University of Copenhagen
Capital Region of Denmark
City of Copenhagen

Richard Petersens Plads
Building 324
2800 Kongens Lyngby, Denmark
cachet@dtu.dk
www.cachet.dk

Summary

This technical report describes the CACHET Unified Methodology for Assessment of Clinical Feasibility (CUMACF). The purpose of this methodology is to provide a standardized way across CACHET projects to assess 'clinical feasibility' of the technologies that are being designed and tested. This report does three things. First it outlines different approaches to establishing 'feasibility' or to run 'pilot studies', including assessment of usability, usefulness, feasibility, efficacy, efficiency, etc. in the development of health technology. Second, it presents CUMACF, including a definition of 'clinical feasibility' and how to assess it. This includes providing concrete guidelines and example of how to follow the methodology and how to set up assessment of 'clinical feasibility'. Finally, the report outlines how to analyze and present data from the method.

Preface

This technical report was prepared at the Copenhagen Center for Health Technology (CACHET)[1]. It is intended to work as a standardized method across research studies done as part of CACHET and to serve as an inspiration for cross-disciplinary research into the design of health technology for the benefit of patients and society in general.

At the time of writing, this is 'work in progress' and the report has not (yet) been published. Hence, any comments, inputs, and correction are most welcome and should be directed to the author.

The report should be cited as:

- Jakob E. Bardram. *CACHET Unified Methodology for Assessment of Clinical Feasibility.* Technical report. Copenhagen, Denmark: Copenhagen Center for Health Technology, 2018. Available from `http://www.cachet.dk/research/cumacf`.

using the following bibtex entry:

```
@techreport{cumacf2018,
  author = {Bardram, Jakob E.},
  title = {{CACHET Unified Methodology for Assessment of Clinical
      Feasibility}},
  institution = {Copenhagen Center for Health Technology},
  address = {Copenhagen, Denmark},
  note = {Available from http://www.cachet.dk/research/cumacf},
  year = {2018}
}
```

.

[1] `www.cachet.dk`

Revision History

Date	Version	Description	Author
4/11-17	0.1	Initial version.	JEB
5/12-17	0.2	Version discussed at workshop	JEB
23/11-18	0.3	Incorporating feedback from workshop participants plus from Julia's use of the method. Version presented and discussed at the CACHET Research Seminar on Nov. 26th 2018.	JEB

Contents

List of Figures

CHAPTER 1

Introduction

Methodologically, the design of health technology, including the growing research into 'Personal Health Technology' [BF16] and 'Digital Health' technologies, sits at the intersection of the design sciences and the clinical sciences. On the one hand, health technologies need to be designed, developed, and refined in a design process, which often relies on modern iterative and user-centered design methodologies. On the other hand, health technologies need to be clinically verified in order to assess clinical safety, efficacy, and effectiveness.

There is a growing awareness that there is a fundamental methodological 'conflict' between these two scientific and methodological positions. From a design perspective, a review of self-care technologies for patients with chronic conditions argues that most studies have "largely privileged a medical perspective" and that there are a number of research 'trends' and 'design-tensions' which can be accommodated by taking a human-computer interaction (HCI) approach [Nun+15]. From a health-oriented perspective, a carefully designed randomized controlled trial (RCT) which minimizes the possibility of bias has become accepted as the 'gold standard' for determining the effectiveness of pharmacological agents, and this approach has been transferred to evaluating non-pharmacological interventions, including health technology. However, the traditional RCT approach has a set of limitations for evaluating health technology, including the fact that the RCT does not permit iterative improvements to the design and that the technology may be outdated by the time the trial is complete [Moh+15].

Despite the different scientific paradigms and methodological approaches between the design sciences and health sciences, there is a growing need to be able to design and develop health technologies while being able to point to health benefits – in particular in the early stages of technology development and evaluation. For this purpose, this technical report introduces the *CACHET Unified Methodology for Assessment of Clinical Feasibility (CUMACF)* methodology. The goal of CUMACF is to help researchers in the process of designing and developing health technology to run what we have chosen to call '*feasibility studies*', i.e. studies that help understand whether the technology under design would be feasible to use in future health interventions, if implemented in full scale.

The purpose of CUMACF is twofold. First, borrowing from a design science perspective, CUMACF seeks to support an iterative design process, with frequent design and evaluation sessions involving real users. The idea is to investigate the feasibility of the technology under design as early as possible – this saves time, effort,

and money. Moreover, in contrast to a traditional RCT in which the intervention (e.g. the medication) is treated as a black box, CUMACF seeks to provide an understanding of the intervention (i.e. the technology under design) by providing insights into which parts of the technology (which features) help achieve a health outcome. Second, borrowing from a health science perspective, CUMACF seeks to investigate health efficacy, i.e. the extent to which an intervention does more good than harm under ideal circumstances [Hay99]. The idea is to investigate potential efficacy during design and as early as possible. CUMACF will not establish a high level of evidence since the study typically does not involve a control group nor has sufficient statistical power. But such a feasibility study will help with understanding the potential of the technology for health efficacy at an early stage and moreover help understand which other parameters, besides the technology itself, need to be (re)designed in order to obtain the desired health outcome.

This technical report is a 'cook book' for CUMACF. Chapter 2 outlines related work in terms of other approaches and methods for running clinical 'feasibility' or 'pilot' studies. Chapter 3 is the main chapter describing the CUMACF method, which focuses on assessing three things; (i) usage adoption, (ii) perceived usefulness and usability, and (iii) health efficacy. Chapter 4 provides guidelines for how to analyze and visualize the data being collected. The appendices provides references to concrete questionnaires and R scripts for doing data analysis and visualizations.

The overall objective of CUMACF is to provide a *standardized* way to assess the 'feasibility' of a health technology during design and development. Such a standardized method will help to compare test results both *within* the iterative design of one specific technology as well as *between* different technologies. The former implies that a design team can assess the progression of their design across multiple iterations of the technology, whereas the latter implies that different technologies – maybe targeting the same health outcome – can be evaluated and compared in a more standardized manner.

CHAPTER 2

Related Work

This chapter presents other approaches to conducting 'feasibility studies' during the design and development of health technology. At the time of writing, this section is not finalized.

2.1 Trials of Intervention Principles: Evaluation Methods for Evolving Behavioral Intervention Technologies

[Moh+15]

2.2 CONSORT-EHEALTH: Improving and Standardizing Evaluation Reports of Web-based and Mobile Health Interventions

[Eys11]

CHAPTER 3

CACHET Unified Methodology for Assessment of Clinical Feasibility (CUMACF)

The CACHET Unified Methodology for Assessment of Clinical Feasibility (CUMACF) is designed to assess early feasibility of a technology-based intervention. It is divided into three parts:

- **Usage Adoption** – the degree to which the patient uses the technology.
- **Perceived Usefulness and Usability** – the likelihood of successful adoption of the technology and acceptance by users.
- **Health Efficacy** – the capacity for beneficial change or therapeutic effect of the intervention provided by the technology.

3.1 Usage Adoption

A core prerequisite for assessing the feasibility of a health technology is to know whether the patient;

- uses the system in the first place
- uses the technology as instructed and prescribed.

To verify this, assessment of *usage adoption*[1] is beneficial. Usage adoption is a relative measure; it assesses to what degree the user uses the technology as compared to

[1]Some call this *'adherence'* to technology use, borrowing the term from clinical studies and treatment. However, we prefer not to use the term 'adherence' since it carries a conotation that

what is expected. For example, if a patient is asked to assess depression level as a daily mood score, the adoption rate is the percentage of days of self-reported mood compared to the number of days the system has been used by the patient. For example, over a 14 week trial of the MONARCA system, a usage adoption of 87% in average was observed [Bar+13].

However, calculating usage adoption depends on a number of factors which have to be taken into account. First, it is important to consider the exact *instructions* the patient is given. Calculating usage adoption only makes sense if the patient has been instructed to use the system according to a specific usage pattern. This can follow a *temporal pattern*, such as on a hourly, daily, or weekly basis; it might be an *event-based pattern*, which instructs the patient to use the system whenever a certain event occur, such as when s/he feels a certain back pain; or it might be *notification-based pattern*, in which the patient gets notified externally to use the system, e.g. by a person running the study or automatically by the system itself. In all three cases, the important part is to be able to establish the *baseline*, i.e. what constitutes 100%. This may be difficult in event-based usage scenarios.

Second, the *length of the study* is important; if usage is expected on a daily basis, then the number of days the system is used constitutes the baseline. However, in a realistic trial, the number of days that each participant actually use the system may vary significantly – for all sorts of reasons, most participants are able to use the system for only parts of the planned study period. For example, there may be delays in handing over equipment and getting it to work, resignation from the study, acute illness and/or hospitalization, etc. Hence, it is important to accurately *define* as well as *record* the 'true' length of a study for each participant. In practice, this may be a hard problem, since this would entail close monitoring of all participants, which scales poorly.

Third, an important factor which often influence the true length of a study is the *availability* of the technology in itself – i.e. system availability – which has to be monitored (and reported) in detail. For example, in the MONARCA study [Bar+13], usage adoption overall was 80% without taking system downtime into account, but 87% when taken into account. Reporting a usage adoption of 87% rather than 80% seems more valid, since the system was not available for users during its downtime. From a software architecture point of view, system availability is, however, dependent on many factors and different system components. For example, consider a classic personal health technology architecture [BF16], which collects data from sensors via a smartphone and relays this back to a backend server. In such a setup, system downtime can occur in different system components, including the local sensor device, the smartphone app, the smartphone itself, the network connection, and all the different components of the backend server. However, whether failures in a system component lead to system failure that influences the ability of the patient to use the system may

the technology is 'prescribed' and 'should' be used like a medical drug. This is often not the way we would like to evaluate technology and normally technology is not 'prescribed'. Therefore, investigating 'adoption' is a better approach and term.

vary quite a lot.

Fourth, if an average overall usage adoption is to be calculated, the *number of participants* are clearly important to know. However, as discussed above, the usage adoption pr. participant depends on a number of factors, and overall usage adoption hence depends on accurate assessment of usage adoption pr. participant.

Based on our experience from several studies – ours and others – has led us to the following *recommendations*:

1. **Instructions** – Always provide exact instructions to each participant as to how often the technology is supposed to be used. We recommend to apply a *temporal* – rather than an event-based – pattern, such as on a hourly, daily, or weekly basis since this makes it more simple to establish the baseline.

2. **Length of study** – Even though participants often are asked to use the technology for a specific period of time (e.g. 6 months), this may in practice vary quite a lot from participant to participant. Hence, it is important to record the exact length of use for *each* participant. Moreover, when reporting usage adoption it if often very valuable to report this *over time* in discrete time periods. For example, if participants are asked to report mood on a daily basis over a period of 6 months, it would be useful to track adoption on a weekly basis to analyze patterns in uptake or abandonment of the technology over time[2].

3. **System availability** – Always log system up- and downtime throughout the study. This has to be done 'seen from the user's point of view'. Hence, system component failures may or may not cause the user to experience problems in using the system. For example, if the system allows for local caching and replication of self-reported daily mood scores (as in the MONARCA system), failure in network connectivity or even downtime in the backend server may not prevent the user in entering his/her daily mood score. On the other hand, if the local app on the smartphone suffers from a bug that prevents the participant from entering data, this will impact usage adoption.

4. **Number of participants** – Always track the number of active participants of the system over time. This is important to calculate exact usage adoption rates, as argued above. Moreover, tracking the numbers of participants over time on a more overall basis also provides a good baseline for the usage of the system.

Taken together, if the above data is collected, detailed statistics on usage adoption can be reported; (i) overall, (ii) over time, and (iii) pr. participant, all of which takes technical issues about system availability into account. In chapter 4 we shall present ways to analyze and present usage adoption statistics.

[2]Tracking the uptake of health technology over time is associated with the study of diffusion of innovation, which in itself is an interesting sub-study [Rog03].

3.2 Perceived Usefulness and Usability

The second part of CUMACF is to measure *perceived usefulness and usability*. According to research into psychometric assessment of technology acceptance, there is a strong correlation between users' perceived usefulness and usability of a system and the likelihood of future successful adoption and acceptance of the technology [Dav89]. For example, a study by Lazar et al. [Laz+15] showed that 80% of all activity tracking devices were abandoned after 2 months. This was mainly due to the fact that "participants perceived the data collected as not useful". Hence, perceived usefulness is key for technology adoption.

CUMACF mainly follows the Unified Theory of Acceptance and Use of Technology (UTAUT) methodology [Ven+03] combined with a few usability questions from the Post-Study System Usability Questionnaire (PSSUQ) scale [Lew92] and two questions related to behavior change from the Behavior Change Wheel (BCW) methodology [MSW11]. By following the UTAUT methodology, which again builds on the Technology Acceptance Model (TAM) methodology, CUMACF is designed to assess the user's intention for *future* acceptance of the technology. As such, CUMACF does not assess usability of *past* usage, which is the case in other usability methods like PSSUQ and SUS[3].

3.2.1 CUMACF Questionnaire

Following the UTAUT methodology, the CUMACF questionnaire consists of five sections and suggests in total 27 questions. The following sections outline the five sections of the CUMACF questionnaire and Table 3.1 to 3.5 list the questions within each section. Each question addresses a specific issue ('target'), is fitted for a specific study participant ('who')[4], and originates from a specific psychometric methodology ('source'). CUMACF applies a 5-point Likert scale, in which the respondent indicates the degree of agreement or disagreement with the statement from 'Strongly disagree' (left) to 'Strongly agree' (right). Figure A.1 in Appendix A shows the entire CUMACF questionnaire.

The CUMACF questionnaire can collect data on the likelihood of successful adoption of technology, acceptance of users, and intentions to use it. In chapter 4 we shall present ways to analyze and present statistics on perceived usefulness and usability.

3.2.1.1 Health Expectancy (HE)

This part of the questionnaire is directed to assess the degree to which an individual believes that using the system will help him/her to attain *gains in health*[5] – also

[3]For reference, the SUS questionnaire is included in Appendix D, the PSSUQ questionnaire is included in Appendix C, and the UTAUT questionnaire is included in Appendix **??**

[4]P: Patient; C: Clinician; A: All.

[5]In the original UTAUT methodology, this category is called 'Performance Expectancy' and lists questions targeting job performance in a working environment. As part of adapting the UTAUT

termed *perceived usefulness*. The questions in this category are related to; overall usefulness, use of the technology, behavior change, health outcome, efficiency in use, productivity, quality of intervention, and reduction of adverse events.

Table 3.1: Health Expectancy (HE) Questions..

ID	Target	Statement	Who	Source
HE1	Usefulness	Overall, I would find the system useful for [handling | improving] my [condition]	P	UTAUT
HE2	Adherence	I would use [system name] as often as instructed [(i.e. on a [daily | weekly | monthly | ...] basis)]	A	BCW
HE3	Behavior	Using [system name] would help me [reduce | increase] my [weight | physical activity | medicine adherence | ...]	P	BCW
HE4	Health	Using [system name] would help me reach my health goals of [reducing my blood pressure | managing my blood sugar and diabetes | reduce depressive symptoms | ...]	P	
HE5	Efficiency	Using [system name] would enable me to achieve my health goals more quickly and more efficiently	P	UTAUT
HE6	Productivity	Using [system name] would increases my productivity in terms of [consulting with more patients | handling more patient cases | ...].	C	UTAUT
HE7	Quality	Using [system name] would increases the quality of [treatment | care | communication b/w me and my doctor | ...]	A	UTAUT
HE8	Safety	Using [system name] would reduce [adverse events such as] [medication errors | readmission to hospital | mis-communication with my doctor | ...]	A	

In question HE2, specific expected usage frequency should be specified. In question HE3, specific behavior change should be specified. This question may be repeated for each behavior if more than one. In question HE3, specific health goal should be specified. Repeat this question for each goal, if more than one. In question HE5-HE7, terms like 'efficiency', 'productivity', and 'quality' are often related to clinical professionals but can be used to ask patients as well, if properly rephrased. In question HE8, specific adverse events can be specified. However, try to avoid the term 'adverse event' as such since it is a rather clinical and technical term. Name specific examples of events.

3.2.1.2 Effort Expectancy (EE)

This part is directed at assessing the degree to which an individual believes that *ease* is associated with use of system – also called *perceived usability* of the technology.

methodology to investigate health issues, this category has been renamed to 'Health Expectancy'.

The questions in this category are associated with general usability issues like; overall usability, understandability, learnability, easy of use, skillfulness, information quality, interface quality, pleasure, and feature fit.

Table 3.2: Effort Expectancy (EE) Questions..

ID	Target	Statement	Who	Source
EE1	Usability	Overall, I would be satisfied with how easy it is to use [system name]	A	PSSUQ
EE2	Understandability	My interaction with [system name] would be clear and understandable.	A	UTAUT
EE3	Learning	It would be easy for me to learn to use [system name]	A	UTAUT
EE4	Easy	I would find [system name] easy to use	A	UTAUT
EE5	Skillful	I would be skillful at using [system name]	A	UTAUT
EE6	Information Quality	The information (such as [error messages | online help | messages | guidelines | tutorials | ...]) provided with [system name] are clear and useful	A	PSSUQ
EE7	Interface Quality	The interface would be effective in helping me complete the [tasks | self-assessment | ...]	A	PSSUQ
EE8	Pleasure	[system name] would be pleasant to use.	A	PSSUQ
EE9	Features	[system name] would have all the [features | functionality | capabilities] that I expect it to have.	A	PSSUQ

3.2.1.3 Social Influence (SI)

This part is directed at assessing the degree to which an individual perceives that *important others* believe s/he should use the system. These questions focus on asking about the influence of each 'important other', including; health professionals, relative, friends and peers, and society in general.

Table 3.3: Social Influence (SI) Questions..

ID	Target	Statement	Who	Source
SI1	Health professionals	My [doctor | psychiatrist | psychologist | nurse | ...] think that I should use [system name].	P	UTAUT
SI2	Relatives	My family [spouse | children | parents | ...] think that I should use [system name].	P	UTAUT
SI3	Friends & Peers	My peer(s) ([friends | colleagues | care community |...]) think that I should use [system name].	P	UTAUT
SI4	Society	As a [Danish] citizen, I am expected to use [system name].	P	UTAUT

3.2.1.4 Facilitating Conditions (FC)

These question query the degree to which an individual believes that an *organizational and technical infrastructure* exists to support use of the system. The questions listed here are related to; specific technical resources, the user's knowledge, and the available support and assistance.

Table 3.4: Facilitating Conditions (FC) Questions..

ID	Target	Statement	Who	Source
FC1	Resources	I would have the resources necessary to use [system name] (such as [laptop | smartphone | ...]).	A	UTAUT
FC2	Knowledge	I would have the knowledge necessary to use [system name].	A	UTAUT
FC3	Support	A specific person (or group) would be available for [assistance | support] with system [difficulties | questions | technical issues].	A	UTAUT

3.2.1.5 Behavioural Intention (BI)

These final questions are targeted at investigating the degree to which an individual *intends to use the system*. These three questions are taken directly from the UTAUT questionnaire and are rather broad and generic in nature. They try to assess as to whether the user intends to use the system, predicts if s/he will use the system, and to what degree s/he has specific plans to start using the system.

Table 3.5: Behavioural Intention (BI) Questions..

ID	Target	Statement	Who	Source
BI1	Intent	I intend to use [system name] in the next [2 | 6 | 12] months.	A	UTAUT
BI2	Predict	I predict I would use [system name] in the next [2 | 6 | 12] months.	A	UTAUT
BI3	Plan	I plan to use [system name] in the next [2 | 6 | 12] months.	A	UTAUT

3.2.2 Adapting the CUMACF Questionnaire

The CUMACF questionnaire as outlined in Figure A.1 contains the minimum set of question relevant for investigating the likelihood of success for the technology introduction and acceptance of users. It can, however, be adapted and tailored to a specific case in a number of ways.

First, the questions should be tailored according to the type of targeted respondents. For example, there is a big difference as to whether the questionnaire is targeting patients or health professionals. In particular, the original UTAUT methodology

was targeting professional use of technology in a work setting and therefore includes questions related to 'efficiency' (HE5) and 'productivity' (HE6). These questions may be less relevant when targeting patients, or might at least need to be reformulated.

Second, the formulation of the questions needs to be adapted to the specific evaluation case. The questions need to be specific to the case – generic questions from e.g. UTAUT or PSSUQ can be very hard for participants to relate to. Therefore, replace all questions with the relevant system name (in the [system name] placeholders), put in the real name for the kind of intervention, health outcome, behavior change, and systems feature wherever relevant. In Figure A.1 all text that needs to be edited is marked with '[...]' and optional text is separated with a '|'.

Third, the questions related to behavior (HE3) and health (HE4) should be repeated, if more than one behavior change or health outcome is part of the intervention being assessed. Remember to be specific about what behavior change the question is about – following the BCW approach, you should rather ask about whether the respondent was able to 'wash hands' (specific behavior) rather than 'increase hygiene' (the intended goal). Similarly with health outcome questions; questions should e.g. ask about 'lower blood pressure' rather than 'manage hypertension'.

Fourth, the questionnaire in Figure A.1 lists very generic questions in the effort expectancy category, each targeting generic usability issues like understandability, learnability, easy of use, etc. However, if you want to investigate specific systems features, then questions that assess such features can be added to the questionnaire. For example, if the system uses a data visualization designed as a butterfly, which is designed to provide feedback about smoking cessation, then a question like "The butterfly would be effective in helping me reduce smoking" could be relevant.

Fifth, questions can be tailored and added to the social influence, facilitating conditions, and behavioral intention category as needed. For example, it might be relevant to ask about the influence of a specific type of person, or there might be some specific facilitating conditions – like access to outpatient psycho-education – which are important to assess.

In general, when you reformulate or add questions to the questionnaire, please note that questions should be formulated in a 'future tense', especially when asking about *perceived* usefulness (health expectancy) and *perceived* usability (effort expectancy). Since CUMACF builds on UTAUT, questions should be designed to assess the user's intention for future acceptance of the technology (i.e. the feasibility of the technology in future use). As such, CUMACF does not assess usability of *past usage* but of *future adoption*. Therefore, most questions should be formulated along the lines of 'would' and 'should'.

Finally, please keep in mind the usability of the questionnaire when adapting it. It may be tempting to add a number of detailed questions related to detailed behavior change, health outcome, and/or different features of the system. However, answering long and detailed questionnaires is a time consuming and tedious task for users. If the survey is short, there is a greater chance that more respondents will complete it.

3.2.3 Instructions

An adapted version of the following text should be used as an instruction to the user on how to fill in the questionnaire.

> The following questionnaire gives you an opportunity to express your satisfaction with the usefulness and usability of [system name]. Your response will help us understand what aspects of the system you are particularly concerned about and the aspects that satisfy you.
>
> While you answer the questions, think about all the ways that you have used [system name]. Note that the questions are in a future tense – i.e. we are interested in your opinion on how you think the system would help improve health if used daily by you going forward, based on your experience of using [system name] until now.
>
> Please read each statement and indicate how strongly you agree or disagree with the statement by clicking a number on the scale.
>
> [Whenever it is appropriate, please write comments to explain your answers.]
>
> Thank you!

3.2.4 Questionnaire or Interview

CUMACF is intended to be used also for more small studies which apply more qualitative methods, such as interviews. During interviews, the CUMACF questionnaire outlined above can be used as an interview guide. By doing so, questions will address all five domains, including health, effort, social influence, facilities, and behavior intentions.

3.3 Health Efficacy

In health intervention studies, a distinction is often made between efficacy, effectiveness, and efficiency [Hay99]:

- *Efficacy* is the extent to which an intervention does more good than harm under ideal circumstances (“Can it work?”).
- *Effectiveness* assesses whether an intervention does more good than harm when provided under usual circumstances of healthcare practice (“Does it work in practice?”).
- *Efficiency measures* the effect of an intervention in relation to the resources it consumes (“Is it worth it?”).

Since the CUMACF focuses on ‘feasibility’, the methodology focuses on establishing *health efficacy*. As pointed out by Haynes [Hay99]:

> "Efficacy trials typically select patients who are carefully diagnosed; are at highest risk of adverse outcomes from the disease in question; lack other serious illnesses; and are most likely to follow and respond to the treatment of interest. This treatment will be prescribed by doctors who are most likely to follow careful protocol; the comparison will be a placebo, not the current best alternative therapy; and participants will receive special attention from staff who supplement or replace those employed in usual clinical settings." [p. 652]

It makes much sense to focus on efficacy during the initial testing of health technology; if the intervention supported by the technology does not work under these ideal conditions it surely will not work under usual conditions.

3.3.1 Quality of Evidence

Health efficacy can be established in many ways, and the chosen approach depends on the desired level of *quality* in the study. In evidence-based medicine (EBM), the following grading systems for assessing the quality of evidence is often used[6]:

- Level I: Evidence obtained from at least one properly designed RCT.
- Level II-1: Evidence obtained from well-designed controlled trials without randomization.
- Level II-2: Evidence obtained from well-designed cohort studies or case-control studies, preferably from more than one center or research group.
- Level II-3: Evidence obtained from multiple time series designs with or without the intervention. Dramatic results in uncontrolled trials might also be regarded as this type of evidence.
- Level III: Opinions of respected authorities, based on clinical experience, descriptive studies, or reports of expert committees.

The best evidence comes from running a RCT in which the intervention supported by the technology is compared to a control group. However, a RCT may not be well-suited for establishing clinical feasibility of a technology in an early stage of development for a number of reasons. First, designing and executing a RCT takes a lot of resources and is very costly. Therefore, if early feasibility is the goal, the cost-benefit ratio of using a RCT is problematic. Second, a RCT assumes that the intervention, including the technology, is kept stable during the RCT period, which is often years. This is in conflict with a need for continuous improvement and rapid release cycles typically done in technology development. Third, a RCT treats the

[6] This is adopted from wikipedia [Wikb] and based on the U.S. Preventive Services Task Force (USPSTF) 1989 guidelines [For89]

intervention including the technology as a 'black box'. The researchers are blinded as to which patients actively receive the 'treatment' supported by the technology and which are in the control group, and the evaluation of the intervention is typically measured in terms of health outcomes. This methodology provides very little insight into the 'inner workings' of the intervention; it does not investigate how well the different components of an intervention – including the technology – actually work. For example, can the patients figure out how to use the smartphone app (usability); do they find it useful; what is the role and benefit from an online chat forum with trained nurses, etc. In short; a RCT is well-suited for establishing clinical evidence for the efficacy of an intervention, but is very badly suited for providing insight into the usefulness, usability, and benefits of the intervention, and its different components, including the technology.

For these reasons, we *do not recommend* setting up a RCT for establishing health efficacy when trying to establish clinical feasibility in an early stage of technology design and development. Instead, we recommend to run a 'Clinical Proof-of-Concept' study [Bar08], which is defined as:

> "The construction of working prototypes of the necessary functionality and infrastructure in sufficient quality to investigate evidence for improving health in daily use for a suitable period of time; a limited but relevant set of people serving as subjects." [p. 184]

As discussed in chapter 2, several studies in the Journal of Medical Internet Research (JMIR) are labelled as a 'pilot study' or 'feasibility study'. Common to these, is that these include a limited number of participants (4–60) over a limited time duration (few days to few months) and do not include any randomization or control group.

In the design of a pilot study, there are a few parameters to consider. First, the *number* of participants (N). In a RCT, N depends on a power calculation. However, many HCI studies typically employ 20 participants [Hor+13] and a review of psychology experiments also recommended 20 persons per condition [SNS11].

Second, in terms of *recruitment*, Haynes [Hay99] points out that efficacy trials typically select patients who are carefully diagnosed; are at highest risk of adverse outcomes from the disease in question; lack other serious illnesses; and are most likely to follow and respond to the treatment of interest. When testing technology, the latter should not be under-estimated; it is important that the patient is motivated and able to use the technology in question, even though it might be early in the design and development and hence unstable. According to the theory of technology diffusion, they must be 'early adopters' [Rog03].

Third, the *duration* of the study depends on the expected time for the intervention to take effect. Some interventions may take effect within a week (e.g. increased physical activity), others over months (e.g. weight loss), or even years (e.g. recovery from major depression). However, a pilot trial is not suited for long-term evaluation

and most feasibility or pilot studies seldom exceed more than 6 month duration of the intervention pr. patient.

Fourth, different norms for *compensation* exists. Monetary compensation is quite common in US-based studies, whereas in Denmark compensation is kept to an absolute minimum, and is intended to only cover direct cost for the patients to participate[7]. A common strategy when evaluating a technology-based intervention is to provide the technology (e.g. a smartphone) including the infrastructure (e.g. subscription) for free (and allow private use). Another common strategy is to give participants a gift voucher after completing the study, but to *not* inform about this in advance [AM04]. Compensation rates often needs to be approved as part of the ethical approval of the study.

Fifth, during the trial it might be necessary to *adapt* the intervention, including the technology, based on observations and findings during the execution. Here it is important to consider to what degree the adaptation is influencing the study. Some adaptations – like fixing bugs in the software or replacing a study nurse – might have very limited influence on the study and are made to ensure that the study runs as planned. Other adaptations – such as replacing the entire user interface of an app or introducing a study nurse to helps patients online – might have significant influence on the outcome of the study. In general, an adaptive clinical trial is a clinical trial that evaluates a medical device or treatment by observing participant outcomes (and possibly other measures, such as side-effects) on a prescribed schedule, and modifying parameters of the trial protocol in accord with those observations [Wika]. The important part is that the trial protocol is set before the trial begins; the protocol pre-specifies the adaptation schedule and processes.

In summary, we recommend to establish initial health efficacy by running a clinical pilot study along the following lines:

- **N** – The number of participants (N) should be minimum 20.
- **Recruitment** – Recruit patients who are carefully diagnosed and who potentially can benefit from the intervention (are significantly ill), while at the same time are early adopters, i.e. have the skills, motivation, and ability to use the technology.
- **Duration** – The intervention pr. patient should not extend more than 6 months.
- **Compensation** – Compensation should be tailored to local ethics guidelines, but often the technology and the infrastructure is provided free of charge.
- **Adaptation** – The study protocol should allow for adaptation during the study. However, this should be restricted to adaptation which only has a limited effect

[7] Af komitelovens § 20, stk. 1, nr. 3 fremgår, at det er en betingelse for at tillade et forskningsprojekt, at: "eventuelt vederlag eller anden ydelse til forsøgspersonerne for deltagelse i et sundhedsvidenskabeligt forskningsprojekt ikke er egnet til at påvirke samtykkeafgivelsen". Vederlag eller andre ydelser til forsøgspersoner i forbindelse med deltagelse i forsøg må ikke have karakter af at være en betalt arbejdsindsats [Kom11].

on the objective and outcome measure of the study, and should primarily be addressing technical enhancements of *non-functional* software qualities such as robustness, security, usability, and scalability.

Once – but not before – a clinical proof-of-concept shows promising results in such a small ($N > 20$) non-randomized trial without a control group, we recommend to move on to plan, design, and execute a RCT.

3.3.2 Outcome Measures

Defining outcome measures is clearly dependent on the health topic in question and the type of disease being addressed. However, a basic distinction can be made according to; (i) *how* health outcome measures are obtained versus (ii) *who* measure it. This distinction is illustrated in Table 3.6. The main categories are:

Measurement – Health outcome measured by a (medical) device. Examples include blood glucose measurement by a glucose meter, blood pressure measurement by an automatic cuff-based blood pressure monitor, and weight measurement by a scale.

- **Clinical measurement** – Measurements based on a medical device done by health professionals. This typically takes place in a clinical setting, by a clinician who has received professional training, who uses clinical-grade equipment, which is a CE-marked or FDA approved medical device. Such measurements are documented by the clinician; typically in a medical record or a dedicated study protocol.
- **Self-measurement** – Measurement done by the patient him- or herself, or non-professional caregivers such as a spouse or parent. This typically takes place in a non-clinical setting like the home of the patient, the patient may have received some basic training, is using a non-certified consumer device, and is responsible for documentation.
- **Automatic measurement** – Measurement done automatically via an autonomous health monitor. This allow for continuous monitoring and data logging without any intervention from human actors (clinicians or patients), but often, however, requires support for mounting and maintaining the device. This approach can apply both medical devices (such as continuous glucose monitors) as well as consumer devices (such as the activity monitor in a smartphone).

Assessment – Health outcome assessed by a human. Examples include diagnosis of depression by a psychiatrist based on a Schedules for Clinical Assessment in Neuropsychiatry (SCAN) interview, self-assessment of depression by the patient by the Patient Health Questionnaire (PHQ), or self-reporting of alcohol intake in a smartphone app.

Clinical assessment – Assessment done by health professionals. This typically takes place in a clinical setting, by a clinician who has received professional training, using a clinical protocol or questionnaire, which has been professionally validated and approved. The detailed assessment follows a strict protocol, and is interpreted and documented by the clinician.

Self-assessment – Assessment done by the patient him- or herself, or non-professional caregivers. This typically takes place in a home setting, the patient has received limited training, and is using a patient-centric questionnaire. Patient-centric assessment methods are often (or should be) verified against clinical assessments.

Automatic assessment – Assessment done automatically via autonomous health monitoring. Based on continuous monitoring of a patient's biomedical and behavioral data, the patient state of health may be automatically assessed and logged. For example, automatic assessment of the level of depression based on behavioral monitoring. This often requires the assessment method (i.e. the algorithm) to be calibrated, trained, and personalized for each patient. And like self-assessment methods, automatic assessment methods are – or should be – verified against clinical assessments methods[8].

Table 3.6 provides an overview of how health outcome can be measured as part of a feasibility study. Methods might be combined and may overlap several categories. For example, studying the feasibility of a hypertension monitoring system may include both the patient to put on and start a cuff-based monitor (self-measurement), as well as self-assessment of sleep, alcohol intake, and physical activity, while also relying on the patient to have clinical measurements of blood pressure, pulse, heart rate variability (HRV), etc. at regular visits to a clinic.

[8] Note that automatic assessment is at the very forefront of research, and few – if any – automatic assessment methods are yet clinically validated and available for general use.

Table 3.6: Taxonomy for measuring health outcome in a clinical pilot study.

	Health Professional	**Patient**	**Automatic**
Measurement (device)	*Clinical measurement* • in clinic • by clinician • approved medical device • documented in medical record	*Self-measurement* • at home by patient • limited training • non-approved medical device • documented by patient	*Automatic measurement* • continuously • mounting and maintenance by patient • any devices • automatic logging
Assessment (human)	*Clinical assessment* • in clinic • trained • verified clinical assessment methods • documented in medical record	*Self-assessment* • at home by patient • limited training • patient reported outcome (PRO) • documented by patient	*Automatic assessment* • continuously • calibration & training • verified against clinical assessment methods • automatic logging

CHAPTER 4

Analysis of Data

This chapter provides some overall guidelines on how to analyze and present data collected from a CUMACF study.

4.1 Usage Adoption Data

Table 4.1 shows a fictive data set for usage adoption. All number are days. As explained in chapter 3, the following data for each participant is important to collect:

- The instructed baseline, i.e. how long the participant is asked to use the system. In Table 4.1, most participants are asked to use the system for 6 month (183 days), but some only one or two month (30/61 days).
- The length of the study for each participant, which – as illustrated – may vary quite significantly from the instructed period.
- The downtime of the technology, as seen from the participants point-of-view. In this case; how many days the system was – for one reason or another – unavailable for each participant.
- Observed usage, i.e. how many days each participant used the system as prescribed.

Using the following simple formula, adoption pr. participant can be calculated, as shown in the last column of Table 4.1:

$$adoption = \frac{usage}{length - downtime}$$

Note that the *instructed* number of days are not included in the calculation of adoption. However, if the actual length of study for each participant is unavailable, the instructed length may substitute this. This will, off course, provide a lower adoption rate. Note also, that the *total adoption* is calculated using the formula above – in this case it is 92%. Note that using the average of each participant's average adoption rate is misleading as the overall adoption rate This is illustrated in Table 4.1, where the average adoption rate is 87%. This is lower, since the adoption rate for the 'short' studies (P7 and P8) are low.

Table 4.1: Example of usage adoption data collected. In this example, all reported number are days of a study..

participant	instr.	lenght	downtime	usage	adoption
P1	183	170	3	165	99%
P2	183	120	2	101	86%
P3	183	73	2	45	63%
P4	183	173	1	156	91%
P5	183	108	1	105	98%
P6	122	93	1	91	99%
P7	61	45	2	23	53%
P8	30	23	0	20	87%
P9	183	194	1	191	99%
P10	183	118	3	115	100%
total		1.117	16	1.012	**92%**
avg.					**87%**

As also argued in Chapter 3, reporting usage adoption over time provides a good insight in the uptake and 'diffusion' of the technology being tested. Figure 4.1 provides an example of how to illustrate usage adoption over time. In this example, the usage from the ten participants listed in Table 4.1 is shown on a monthly basis over a period of 15 month (month 3 to 17). The top figure shows the usage patterns for each of the ten participants with a smoothed curve fitted to the data points. Participants show different usage patterns. For example, P2 initially starts using the system, but ends up with limited use of the system, whereas P10 starts out low, but gradually increase her/his usage. The bottom figure shows the overall usage. This latter figure can illustrate the overall diffusion of the technology. According to the theory of diffusion of technology (innovation), this should be a normal distribution over time [Rog03]. This trend is recognized in Figure 4.1; usage gradually grows from month 3, raising to a plateau in months 7 to 12, after which it declines. This patterns is, of course, contingent to the specific details of the study; in our example, the study period is 6 months and participants did not use the system beyond this period. Figure 4.1 is generated from an R script, which is available in Appendix E.

4.2 Perceived Usefulness and Usability Data

The CUMACF questionnaire is applying a 5-point Likert scale of; 'strongly disagree', 'disagree', 'neither agree nor disagree', 'agree', and 'strongly agree', with numerical scores from 1–5. The question is how to represent the results of a survey using such a 5-point Likert. One common practice is to take the mean. However, as pointed out by Robbins et al. [R+11], it is controversial since there is no assurance that there is even spacing between the descriptions of attitude. There is no reason to assume that the distance between agree and strongly agree is the same as the distance from agree to neither agree nor disagree. However, even if it were acceptable to take means, it is not very useful. For example, if we look at the example survey data in Table 4.2, the first three questions (HE1–3) provides the same mean (24.0), but there is a big

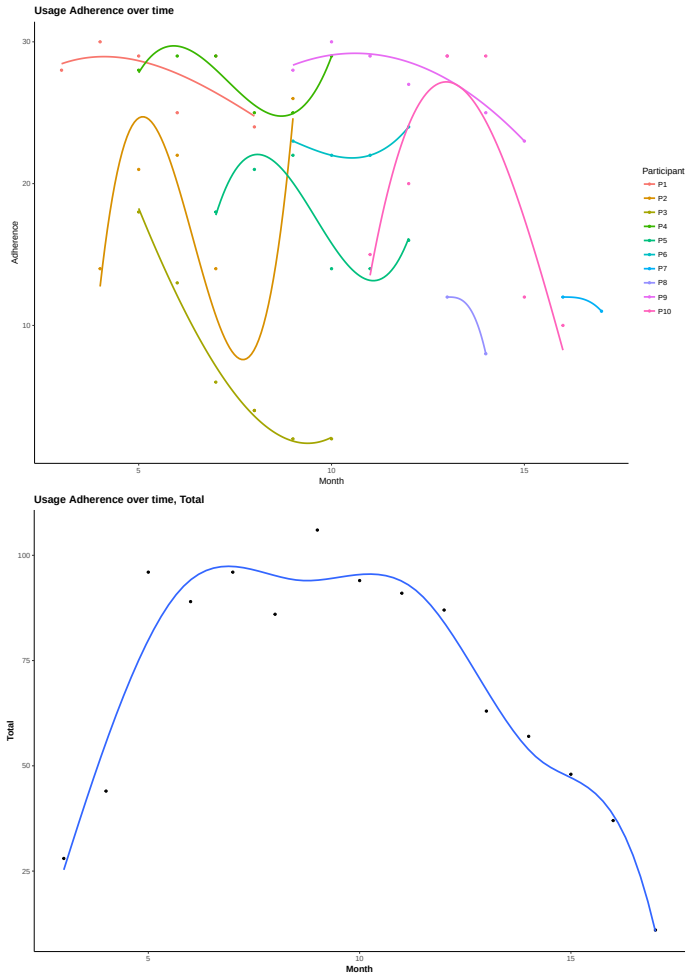

Figure 4.1: Usage of a system over time. Top: Usage patterns for each of the ten participants. Bottom: Total usage pattern.

difference between HE1 where respondents are concentrated at both ends of the scale, and HE2 in which all respondents are all neutral. Hence, based on the response to HE1 it would be very wrong to conclude that "on average, respondents were *neutral* as to whether the system would be useful for handling diabetes".

Robbins et al. [R+11] discuss different ways to present and visualize Likert scale data and recommend to present data in (i) a table and (ii) as a so-called 'diverging stacked bar chart' As an example, we can look at the data in Table 4.2, which is visualized in a diverging stacked bar chart in Figure 4.2. Figure 4.2 is generated from

Table 4.2: Example of survey data from the CUMACF perceived usefulness and usability questionnaire. The center figures are the number of respondents in each category, and total and average scores are on the right..

#	Question	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	Scores Total	Scores Avg.
HE1	Usefulness	20	0	0	0	20	120	24.0
HE2	Adoption	0	0	40	0	0	120	24.0
HE3	Behavior	10	10	0	10	10	120	24.0
HE4	Health	12	2	4	6	23	167	33.4
HE5	Efficiency	2	14	32	21	3	225	45,0
HE6	Productivity	32	2	3	12	2	103	20.6
HE7	Quality	10	2	4	1	23	145	29.0
HE8	Safety	4	14	2	33	3	185	37.0
EE1	Usability	12	2	5	2	3	54	10.8
EE2	Understandable	10	2	4	6	23	165	33.0
EE3	Learning	4	2	23	12	11	180	36.0
EE4	Easy	28	11	5	4	3	96	19.2
EE5	Skillful	18	2	4	6	11	113	22.6
EE6	Information	4	14	32	15	3	203	40.6
EE7	Interface	5	21	5	4	3	92	18.6
EE8	Pleasure	12	14	11	3	4	105	21.0
EE9	Features	4	4	3	44	12	257	51.4

an R script (originally proposed by Heiberger & Robbins [HR14]). The R script is available in Appendix E.

4.3 Health Outcome

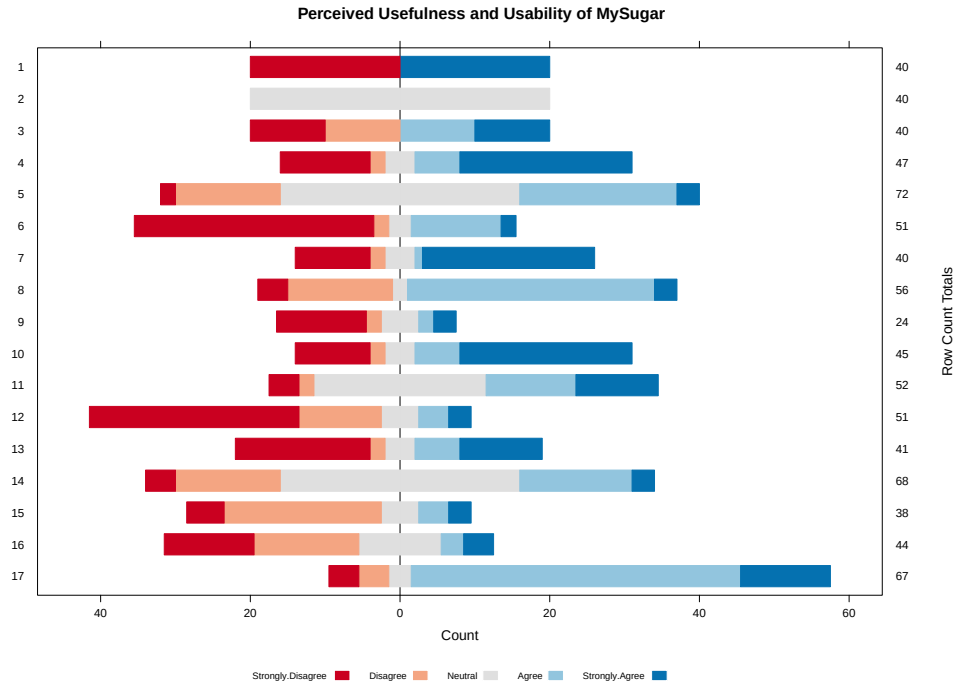

Figure 4.2: Diverging stacked bar chart of the data in Table 4.2.

APPENDIX A

The CACHET Unified Methodology for Assessment of Clinical Feasibility (CUMACF) Questionnaire

No.	Target	Statement	Who	Source	Note
Health expectancy (HE) : The degree to which an individual believes that using the system will help him to attain gains in job performance					
HE1	Usefulness	Overall, I would find the system useful for [handling | improving] my [condition]	P	UTAUT	This question is related to the overall outcome goal.
HE2	Adherence	I would use [system name] as often as instructed [(i.e. on a [daily | weekly | monthly | …] basis)]	A	BCW	Specify specific expected usage frequency.
HE3	Behavior	Using [system name] would help me [reduce | increase] my [weight | physical activity | medicine adherence | …]	P	BCW	Specify specific behavior change. Repeat this question for each behvior if
HE4	Health	Using [system name] would help me reach my health goals of [reducing my blood pressure | managing my blood sugar and diabetes | reduce depressive sympthoms | …]	P		Specify specific health goal. Repeat this question for each goal, if more than one.
HE5	Efficiency	Using [system name] would enable me to achieve my health goals more quickly and more efficiently	P	UTAUT	
HE6	Productivity	Using [system name] would increases my productivity in terms of [consulting with more patients | handling more patient cases |	C	UTAUT	Efficientcy, productivity, and quality are often related to clinical professionals but can bes used to ask
HE7	Quality	Using [system name] would increases the quality of [treatment | care | communication b/w me and my doctor | …]	A	UTAUT	patients
HE8	Safety / Adverse Events	Using [system name] would reduce [adverse events such as] [medication errors | readmission to hospital | miscommunication with my doctor | …]	A		Specify specific adverse events - and try to avoid the term all together since it is a clinical term. Repeat this question for each event, if more than
Effort Expectancy (EE) : The degree to which an individual believes that ease is associated with use of system					
EE1	Usability	Overall, I would be satisfied with how easy it is to use [system name]	A	PSSUQ	
EE2	Understandable	My interaction with [system name] would be clear and understandable.	A	UTAUT	
EE3	Learning	It would be easy for me to learn to use [system name]	A	UTAUT	
EE4	Easy	I would find [system name] easy to use	A	UTAUT	
EE5	Skillful	I would be skillful at using [system name]	A	UTAUT	
EE6	Information Qality	The information (such as [error messages | online help | messages | guidelines | tutorials | …]) provided with [system name] are clear and useful	A	PSSUQ	
EE7	Interface Quality	The interface would be effective in helping me complete the [tasks | self-assessment | …]	A	PSSUQ	
EE8	Pleasure	[system name] would be plesant to use	A	PSSUQ	
EE9	Features	[system name] would have all the [features | functionalities | capabilities] that I expect it to have	A	PSSUQ	
Social Influence (SI) : The degree to which an individual perceived that important others believe s/he should use the system					
SI1	Health professionals	My [doctor | psychiatrist | psychologist | nurse | …] think that I should use [system name].	P	UTAUT	
SI2	Relatives	My family [spouse | children | parents | …] think that I should use [system name].	P	UTAUT	
SI3	Friends & Peers	My peer(s) ([friends | colleagues | care community |…]) think that I should use [system name].	P	UTAUT	
SI4	Society	As a [Danish] citizen, I am expected to use [system name].	P	UTAUT	
Facilitating Conditions (FC) : The degree to which an individual believes that an organizational and technical infrastructure exist to support use of					
FC1	Resources	I would have the resources necessary to use [system name] (such as [laptop | smartphone | …]).	A	UTAUT	
FC2	Knowledge	I would have the knowledge necessary to use [system name].	A	UTAUT	
FC3	Support	A specific person (or group) would be available for [assistance | support] with system [difficulties | questions | technical issues].	A	UTAUT	
Behavioural intention (BI) to use : The degree to which an individual intends to use the system					
BI1	Intent	I intend to use [system name] in the next [2 | 6 | 12] months	A	UTAUT	
BI2	Predict	I predict I would use [system name] in the next [2 | 6 | 12]	A	UTAUT	
BI3	Plan	I plan to use [system name] in the next [2 | 6 | 12] months	A	UTAUT	

Figure A.1: The CACHET Unified Methodology for Assessment of Clinical Feasibility (CUMACF) questionnaire.

APPENDIX B

The Simple Usability Scale (SUS) Questionnaire

The Simple Usability Scale (SUS) was designed at Digital Equipment Corporation (DEC) in 1986 and is a simple, ten-item scale giving a global view of subjective assessments of usability [Bro+96]. It covers a variety of aspects of system usability, such as the need for support, training, and complexity, and thus has a high level of face validity for measuring usability of a system. The SUS scale is generally used after the respondent has had an opportunity to use the system being evaluated, but before any debriefing or discussion takes place.

The so-called *SUS score* yields a single number representing a composite measure of the overall usability of the system being studied. Note that scores for individual items are not meaningful on their own. To calculate the SUS score, first sum the score contributions from each item. Each item's score contribution will range from 0 to 4. For items 1,3,5,7,and 9 the score contribution is the scale position minus 1. For items 2,4,6,8 and 10, the contribution is 5 minus the scale position. Multiply the sum of the scores by 2.5 to obtain the overall SUS score [0–100].

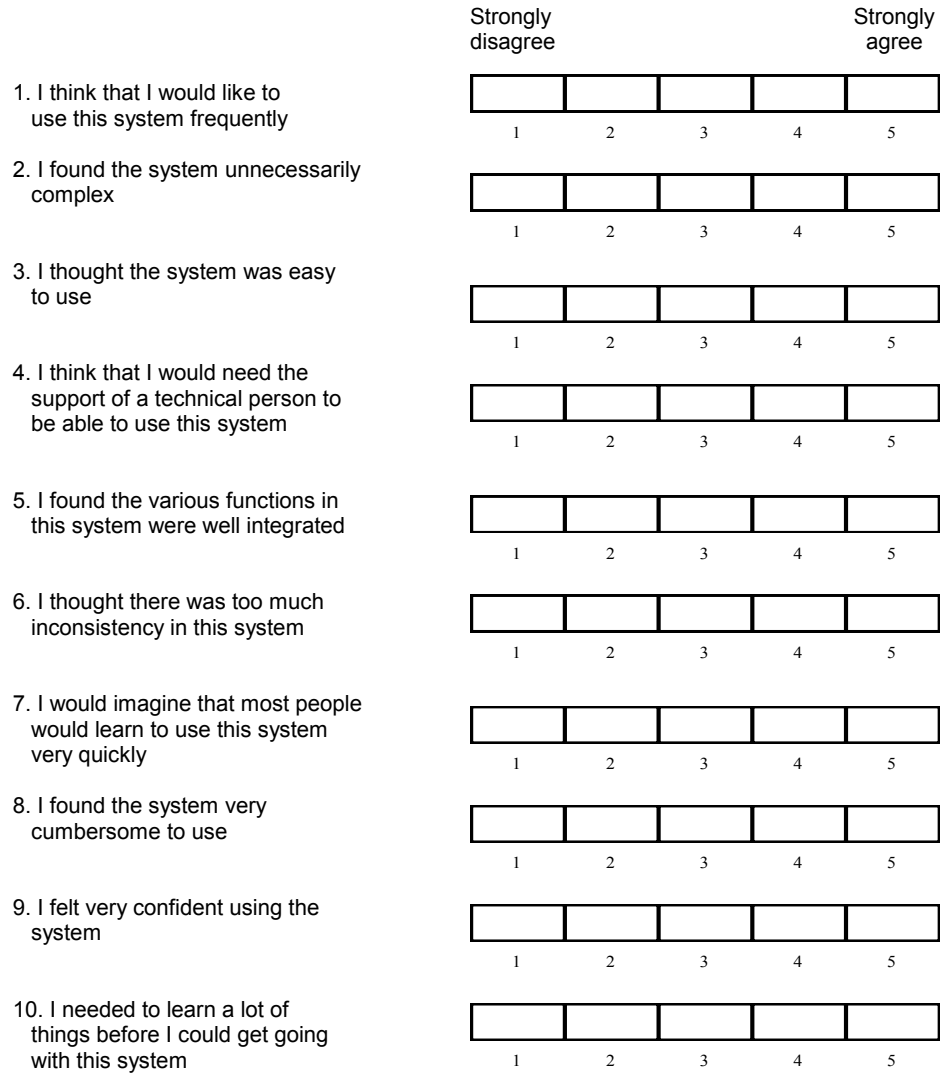

Figure B.1: The Simple Usability Scale (SUS) Questionnaire.

APPENDIX C

The Post-Study System Usability Questionnaire (PSSUQ) Questionnaire

Post-Study System Usability Questionnaire (PSSUQ) consists of 19 items aimed to address the following five system usability characteristics: quick completion of work, ease of learning, high-quality documentation and online information, functional adequacy and rapid acquisition of usability experts and several different user groups [Lew02] that were identified by a panel of IBM HCI experts.

The Post-Study System Usability Questionnaire Items

The first item illustrates the item format. The remaining items show only the item text to conserve space. Each item also has an area for comments (not shown).

1. Overall, I am satisfied with how easy it is to use this system.

STRONGLY AGREE						STRONGLY DISAGREE	
1	2	3	4	5	6	7	N/A

2. It was simple to use this system.
3. I could effectively complete the tasks and scenarios using this system.
4. I was able to complete the tasks and scenarios quickly using this system.
5. I was able to efficiently complete the tasks and scenarios using this system.
6. I felt comfortable using this system.
7. It was easy to learn to use this system.
8. I believe I could become productive quickly using this system.
9. The system gave error messages that clearly told me how to fix problems.
10. Whenever I made a mistake using the system, I could recover easily and quickly.
11. The information (such as on-line help, on-screen messages and other documentation) provided with this system was clear.
12. It was easy to find the information I needed.
13. The information provided for the system was easy to understand.
14. The information was effective in helping me complete the tasks and scenarios.
15. The organization of information on the system screens was clear.

Note: The "interface" includes those items that you use to interact with the system. For example, some components of the interface are the keyboard, the mouse, the microphone, and the screens (including their use of graphics and language).

16. The interface of this system was pleasant.
17. I liked using the interface of this system.
18. This system has all the functions and capabilities I expect it to have.
19. Overall, I am satisfied with this system.

Figure C.1: The Post-Study System Usability Questionnaire (PSSUQ) Items.

APPENDIX D

The Unified Theory of Acceptance and Use of Technology (UTAUT) Questionnaire

The Unified Theory of Acceptance and Use of Technology (UTAUT) [Ven+03] is a an instrument, which is a synthesis of eight existing models of technology acceptance – including the Technology Acceptance Model (TAM). UTAUT also integrates elements from: Theory of Reasoned Action, Motivational Model, Theory of Planned Behaviour (TPB), a combined TAM and TPB model, Model of PC Utilization, Innovation Diffusion Theory, and Social Cognition Theory. The unification of these models provides UTAUT with eight constructs: Performance expectancy, Effort expectancy, Attitude towards using technology, Social influence, Facilitating conditions, Self-efficacy, Anxiety and Behavioural intention to use the system.

Table 16. Items Used in Estimating UTAUT

Performance expectancy

U6:	I would find the system useful in my job.
RA1:	Using the system enables me to accomplish tasks more quickly.
RA5:	Using the system increases my productivity.
OE7:	If I use the system, I will increase my chances of getting a raise.

Effort expectancy

EOU3:	My interaction with the system would be clear and understandable.
EOU5:	It would be easy for me to become skillful at using the system.
EOU6:	I would find the system easy to use.
EU4:	Learning to operate the system is easy for me.

Attitude toward using technology

A1:	Using the system is a bad/good idea.
AF1:	The system makes work more interesting.
AF2:	Working with the system is fun.
Affect1:	I like working with the system.

Social influence

SN1:	People who influence my behavior think that I should use the system.
SN2:	People who are important to me think that I should use the system.
SF2:	The senior management of this business has been helpful in the use of the system.
SF4:	In general, the organization has supported the use of the system.

Facilitating conditions

PBC2:	I have the resources necessary to use the system.
PBC3:	I have the knowledge necessary to use the system.
PBC5:	The system is not compatible with other systems I use.
FC3:	A specific person (or group) is available for assistance with system difficulties.

Self-efficacy

I could complete a job or task using the system…

SE1:	If there was no one around to tell me what to do as I go.
SE4:	If I could call someone for help if I got stuck.
SE6:	If I had a lot of time to complete the job for which the software was provided.
SE7:	If I had just the built-in help facility for assistance.

Anxiety

ANX1:	I feel apprehensive about using the system.
ANX2:	It scares me to think that I could lose a lot of information using the system by hitting the wrong key.
ANX3:	I hesitate to use the system for fear of making mistakes I cannot correct.
ANX4:	The system is somewhat intimidating to me.

Behavioral intention to use the system

BI1:	I intend to use the system in the next <n> months.
BI2:	I predict I would use the system in the next <n> months.
BI3:	I plan to use the system in the next <n> months.

Figure D.1: The Unified Theory of Acceptance and Use of Technology (UTAUT) Questionnaire. From [Ven+03].

APPENDIX E

R Script for Plotting Data

E.1 Plotting Usage Adherence Data

The following matrix shown the raw adherence scores used in the example shown in Table 4.1 and plotted in Figure 4.1.

#	Month	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	Total
1	3	28	NA	NA	NA	NA	NA	NA	NA	NA	NA	28
2	4	30	14	NA	NA	NA	NA	NA	NA	NA	NA	44
3	5	29	21	18	28	NA	NA	NA	NA	NA	NA	96
4	6	25	22	13	29	NA	NA	NA	NA	NA	NA	89
5	7	29	14	6	29	18	NA	NA	NA	NA	NA	96
6	8	24	12	4	25	21	NA	NA	NA	NA	NA	86
7	9	NA	11	2	20	22	23	NA	NA	28	NA	106
8	10	NA	12	2	14	14	22	NA	NA	30	NA	94
9	11	NA	11	NA	NA	14	22	NA	NA	29	15	91
10	12	NA	NA	NA	NA	16	24	NA	NA	27	20	87
11	13	NA	NA	NA	NA	NA	NA	NA	12	29	22	63
12	14	NA	NA	NA	NA	NA	NA	NA	8	25	24	57
13	15	NA	NA	NA	NA	NA	NA	NA	NA	23	25	48
14	16	NA	NA	NA	NA	NA	NA	12	NA	NA	25	37
15	17	NA	NA	NA	NA	NA	NA	11	NA	NA	NA	11

The following R scrips is used to generate the plots in Figure 4.1.

```
# A simple example of plotting fitted curves for usage adherence pr.
    participant and in total
# Jakob E. Bardram, 2017

library(ggplot2)
library(xts)
library(zoo)

#loading adherence data
```

```
adherence <- read.csv("~/Dropbox/WRITINGS/2017.CACHET.User.Study.Methodology/method/adherence.csv", sep=";")
adh_data <- adherence

# stacking the data into three columns [Month, Adherence, Participant] which is to be used by ggplot next
# note that the first and last columns of the adherence data are not included (Month and Total)
col_count <- ncol(adh_data) - 1
adh_frame <- data.frame(adh_data["Month"],stack(data.frame(coredata(adh_data[c(2:col_count)]))))
names(adh_frame) <- c("Month", "Adherence", "Participant")

# creating a theme for the graphs
t <-    theme(panel.background=element_rect(fill = "white"),
              panel.grid.minor = element_blank(),
              panel.grid.major = element_blank(),
              axis.line = element_line(colour = "black", size = 0.3),
              legend.background=element_rect(fill = "white"),
              legend.key=element_rect(fill = "white"),
              title = element_text(lineheight=.8, face="bold")
)

# plotting the data for all participants - showing both points and a smooth 'spline' trend line
plot <- ggplot(adh_frame, aes(x=Month, y=Adherence, color=Participant))
plot <- plot + geom_point(aes(x=Month, y=Adherence, color=Participant), size = 1)
plot <- plot + geom_smooth(method = "lm", formula = y ~ splines::bs(x, 4), se = FALSE)
plot <- plot + ggtitle("Usage Adherence over time")
plot <- plot + t
plot

#stacking the Total column
adh_total <- data.frame(adh_data["Month"],data.frame(adh_data["Total"]))

#plotting the Total adherence over time, smooth
plot2 <- ggplot(adh_total, aes(x=Month, y=Total))
plot2 <- plot2 + geom_point(aes(x=Month, y=Total), size = 1)
plot2 <- plot2 + geom_smooth(method = "lm", formula = y ~ splines::bs(x, 7), se = FALSE)
plot2 <- plot2 + ggtitle("Usage Adherence over time, Total")
plot2 <- plot2 + t
plot2

# a plot of the data as a stacked area chart -- not smoothing, so not so nice...
plot3 <- ggplot(adh_frame, aes(x=Month, y=Adherence, color=Participant))
plot3 <- plot3 +
  geom_area(aes(colour = Participant, fill= Participant), position = 'stack' )
plot3 <- plot3 +
  theme(panel.background=element_rect(fill = "white"),
        panel.grid.minor = element_blank(),
```

```
        panel.grid.major = element_blank(),
        axis.line = element_line(colour = "black", size = 0.3),
        legend.background=element_rect(fill = "white"),
        legend.key=element_rect(fill = "white"),
        plot.title = element_text(lineheight=.8, face="bold")
  )

plot3
```

.

E.2 Generating Diverging Stacked Bar Charts for Likert Scale Data

The R script generating the so-called 'Diverging Stacked Bar Charts' for Likert scales visualization was originally proposed by Heiberger & Robbins [HR14]). The following R script is used to generate Figure 4.2 from the data in Table 4.2 (without the 'Total' and 'Avg.' columns). The script is adopted from a script proposed by ´Wesley' at r-bloggers.com[1].

```
# A simple example of a 'Diverging Stacked Bar Chart' for Likert Scale data
    on perceived usefulness and usability
# Based on example from https://www.r-bloggers.com/plotting-likert-scales/
# Jakob E. Bardram, 2017

require(grid)
require(lattice)
require(latticeExtra)
require(HH)

#loading survey data
sgbar.likert<- survey
title<-"Perceived Usefulness and Usability of MySugar"

# A very simple plot -- out of the box
plot.likert(sgbar.likert, main=title)

# Changing the color palette
pal<-brewer.pal((numlevels-1),"RdBu")
pal[ceiling(numlevels/2)]<-"#DFDFDF"
# A slightly more tailored plot
plot.likert(sgbar.likert,
            main=title,
            col=pal,
            reference.line.col=c('black'),
            strip.left=FALSE,
            rightAxis=TRUE,
            sub="5-point Likert Scale"
```

[1] https://www.r-bloggers.com/plotting-likert-scales/

```
28 )
```

.

APPENDIX F

Acronyms

CACHET Copenhagen Center for Health Technology

EBM evidence-based medicine

RCT randomized controlled trial

JMIR Journal of Medical Internet Research

HCI human-computer interaction

CUMACF CACHET Unified Methodology for Assessment of Clinical Feasibility

UTAUT Unified Theory of Acceptance and Use of Technology

PSSUQ Post-Study System Usability Questionnaire

BCW Behavior Change Wheel

SUS Simple Usability Scale

TAM Technology Acceptance Model

SCAN Schedules for Clinical Assessment in Neuropsychiatry

PHQ Patient Health Questionnaire

HRV heart rate variability

Bibliography

[AM04] Jacob Anhøj and Claus Møldrup. "Feasibility of Collecting Diary Data From Asthma Patients Through Mobile Phones and SMS (Short Message Service): Response Rate Analysis and Focus Group Evaluation From a Pilot Study". In: *Journal of Medical Internet Research (JMIR)* 6.4 (December 2004), e42. DOI: 10.2196/jmir.6.4.e42. URL: http://www.ncbi.nlm.nih.gov/pubmed/15631966.

[Bar+13] Jakob E Bardram et al. "Designing mobile health technology for bipolar disorder: a field trial of the monarca system". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM, 2013. ISBN: 978-1-4503-1899-0. URL: http://doi.acm.org/10.1145/2470654.2481364.

[Bar08] Jakob E Bardram. "Pervasive Healthcare as a Scientific Dicipline". In: *Methods of Information in Medicine* 3.47 (2008), pages 129–142. DOI: http://dx.doi.org/10.3414/ME9107. URL: http://dx.doi.org/10.3414/ME9107.

[BF16] Jakob E Bardram and Mads Frost. "The Personal Health Technology Design Space". In: *IEEE Pervasive Computing* 15.2 (2016), pages 70–78.

[Bro+96] John Brooke et al. "SUS-A quick and dirty usability scale". In: *Usability evaluation in industry* 189.194 (1996), pages 4–7.

[Dav89] Fred D Davis. "Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology". In: *MIS Quarterly* 13.3 (September 1989), pages 319–339.

[Eys11] Gunther Eysenbach. "CONSORT-EHEALTH: Improving and Standardizing Evaluation Reports of Web-based and Mobile Health Interventions". In: *J Med Internet Res* 13.4 (December 2011), e126. ISSN: 1438-8871. DOI: 10.2196/jmir.1923. URL: http://www.ncbi.nlm.nih.gov/pubmed/22209829.

[For89] US Preventive Services Task Force. *Guide to clinical preventive services: Report of the US Preventive Services Task Force*. DIANE Publishing, 1989.

[Hay99] Brian Haynes. “Can it work? Does it work? Is it worth it?: The testing of healthcare interventions is evolving”. In: *BMJ: British Medical Journal* 319.7211 (1999), page 652.

[Hor+13] Kasper Hornbæk et al. “Some whys and hows of experiments in human–computer interaction”. In: *Foundations and Trends® in Human–Computer Interaction* 5.4 (2013), pages 299–373.

[HR14] Richard M Heiberger and Naomi B Robbins. “Design of diverging stacked bar charts for Likert scales and other applications”. In: *Journal of Statistical Software* 57.5 (2014), pages 1–32.

[Kom11] National Videnskabsetisk Komité. *Appendiks 1: Retningslinjer for vederlag eller andre ydelser til frivillige forsøgspersoner*. December 2011. URL: `http://www.nvk.dk/~/media/NVK/Dokumenter/Appendiks-1.pdf`.

[Laz+15] Amanda Lazar et al. “Why We Use and Abandon Smart Devices”. In: *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. UbiComp ’15. Osaka, Japan: ACM, 2015, pages 635–646. ISBN: 978-1-4503-3574-4. DOI: `10.1145/2750858.2804288`. URL: `http://doi.acm.org/10.1145/2750858.2804288`.

[Lew02] James R Lewis. “Psychometric evaluation of the PSSUQ using data from five years of usability studies”. In: *International Journal of Human-Computer Interaction* 14.3-4 (2002), pages 463–488.

[Lew92] James R Lewis. “Psychometric evaluation of the post-study system usability questionnaire: The PSSUQ”. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. Volume 36. 16. SAGE Publications Sage CA: Los Angeles, CA. 1992, pages 1259–1260.

[Moh+15] David C Mohr et al. “Trials of intervention principles: evaluation methods for evolving behavioral intervention technologies”. In: *Journal of medical Internet research* 17.7 (2015).

[MSW11] Susan Michie, Maartje M van Stralen, and Robert West. “The behaviour change wheel: a new method for characterising and designing behaviour change interventions”. In: *Implementation science* 6.1 (2011), page 42.

[Nun+15] Francisco Nunes et al. “Self-care technologies in HCI: Trends, tensions, and opportunities”. In: *ACM Transactions on Computer-Human Interaction (TOCHI)* 22.6 (2015), page 33.

[R+11] Naomi B Robbins, Richard M Heiberger, et al. “Plotting Likert and other rating scales”. In: *Proceedings of the 2011 Joint Statistical Meeting*. 2011, pages 1058–1066.

[Rog03] Everett M. Rogers. *Diffusion of Innovations*. 5th edition. Simon and Schuster, 2003.

[SNS11] Joseph P Simmons, Leif D Nelson, and Uri Simonsohn. “False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant”. In: *Psychological science* 22.11 (2011), pages 1359–1366.

[Ven+03] Viswanath Venkatesh et al. “User acceptance of information technology: Toward a unified view”. In: *MIS quarterly* (2003), pages 425–478.

[Wika] Wikipedia. *Adaptive clinical trial*. [Online; accessed 23-October-2017]. URL: `https://en.wikipedia.org/wiki/Adaptive_clinical_trial`.

[Wikb] Wikipedia. *Evidence-based medicine*. [Online; accessed 23-October-2017]. URL: `https://en.wikipedia.org/wiki/Evidence-based_medicine`.

{2.7182818284}ϑφερτυθιοπσδφγηξκλ