THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Improving quality of summative eHealth evaluations

MONIKA NAIR

Department of Technology Management and Economics

CHALMERS UNIVERSITY OF TECHNOLOGY

Gothenburg, Sweden 2021

IMPROVING QUALITY OF SUMMATIVE EHEALTH EVALUATIONS
MONIKA NAIR
ISBN 978-91-7905-562-2

Doktorsavhandlingar vid Chalmers tekniska högskola
Ny serie nr. 5029
ISSN 0346-718X

Department of Technology Management and Economics
Chalmers University of Technology
SE-412 96 Gothenburg
Sweden
Telephone + 46 (0)31-772 1000

Printed by Chalmers Reproservice
Gothenburg, Sweden 2021

ABSTRACT

Summative evaluation, which is conducted at the end of an eHealth trial or implementation, assesses outcomes, produces evidence, and advances knowledge of eHealth implementations in healthcare provisions. Therefore, its high quality is essential in order to reap the benefits of the results generated by evaluation studies. Quality is achieved in eHealth evaluation through contextual sensitivity, adequate research design, adherence to standards, a mixed-method research approach, and ethical handling of data. However, insufficient quality in eHealth evaluation studies leads to decision makers and other potential users disregarding their results, which leads to the resources and effort involved in conducting an evaluation being wasted.

The purpose of this thesis is to study how summative eHealth evaluations can be improved to support the determination of eHealth value in a specific context as well as the use of evidence produced during eHealth interventions. This thesis is built on a single case study of a summative eHealth evaluation of an eHealth implementation project within different healthcare contexts. The thesis focuses on the different phases of evaluation process, assesses adequacy of standards, explores value that the eHealth intervention delivered in different contexts, and studies how evidence from evaluation is further used.

The thesis extends knowledge on eHealth evaluation quality by providing deeper insights into the problems in the existing quality criteria and by introducing two new criteria for quality in eHealth evaluations: capturing value of an eHealth solution and involving healthcare professionals in the intervention and its evaluation. The thesis reveals that meeting some of the criteria is not always practical, and that evaluators might make trade-offs among the criteria. The findings point to a need to improve methodologies for eHealth evaluations by providing better guidance to evaluators and validating evaluation standards in different locations. The thesis also suggests viewing value of an eHealth solution as a holistic view of the created monetary and nonmonetary benefits of eHealth that require monetary and nonmonetary sacrifices in a particular context. In addition, the thesis proposes a model for assessing value of an eHealth solution.

Keywords: Summative evaluation, eHealth, Quality, Standard, Value, Evidence, Translation, Interorganizational collaboration

List of appended papers

The following list of the appended papers contains my previous last name, Jurkeviciute.

Paper 1: Planning a holistic summative eHealth evaluation in an interdisciplinary and multi-national setting: A case study and propositions for guideline development

Jurkeviciute, M., Enam, A., Torres-Bonilla, J., Eriksson, H. (2021).

Published in *BMC Medical Informatics and Decision Making*, *21*(1), 1–13.

Contributions: Monika Jurkeviciute was the main author, initiated and designed the study, collected data, conducted analysis, and co-wrote the paper. Amia Enam and Johanna Torres-Bonilla participated in the study design, data analysis, and writing of the paper. Henrik Eriksson participated in the study design and writing of the paper.

Paper 2: Standards as applied in reality: A case study on the translation of standards in eHealth evaluation practice

Jurkeviciute, M. (2019).

Published in *BMC Medical Informatics and Decision Making*, *19*(1), 1–9.

Contributions: Monika Jurkeviciute was the sole author of the paper.

Paper 3: An Italian Business Case for an eHealth Platform to Provide Remote Monitoring and Coaching Services for Elderly with Mild Cognitive Impairment and Mild Dementia

Jurkeviciute, M., Van Velsen, L., Trimarchi, P. D., Sarvari, L., & Giunco, F. (2019)

Published in *Conference Proceedings ICT for Aging Well – ICT4AWE,* pp. 28–37.

Contributions: Monika Jurkeviciute participated in study design, data collection and analysis, and co-wrote the paper. Lex van Velsen participated in study design and data analysis and provided scientific advisory. Pietro Davide Trimarchi participated in the data collection and analysis and co-wrote the paper. Ladan Sarvari participated in the data analysis. Fabrizio Giunco participated in the data collection.

Paper 4: Identifying the Value of an eHealth Intervention Aimed at Cognitive Impairments: Observational Study in Different Contexts and Service Models

Jurkeviciute, M., Van Velsen, L., Eriksson, H., Lifvergren, S., Trimarchi, P. D., Andin, U., &

Svensson, J. (2020).

Published in *Journal of Medical Internet Research*, *22*(10), e17720.

Contributions: Monika Jurkeviciute participated in study design, data collection and analysis, and co-wrote the paper. Lex van Velsen participated in study design, data collection and analysis and co-wrote the paper. Henrik Eriksson participated in study design and co-wrote the paper. Pietro Davide Trimarchi, Svante Lifvergren, Ulla Andin and Johan Svensson contributed to the data collection and analysis and co-wrote the paper.

Paper 5: Exploring the Use of Evidence From the Development and Evaluation of an Electronic Health (eHealth) Trial: Case Study

Jurkeviciute, M., & Eriksson, H. (2020).

Published in *Journal of Medical Internet Research*, *22*(8), e17718.

Contributions: Monika Jurkeviciute participated in study design, data collection and analysis, and co-wrote the paper. Henrik Eriksson participated in study design, data collection, and co-wrote the paper.

Academic contribution not included in the thesis

What makes an effective literature seminar? A study from a quality management perspective

Altuntas Vural, C., Jurkeviciute, M., Raharjo, H. (2020).

Presented at KUL 2020: Conference on Teaching and Learning.

Acknowledgements

I would like to express my sincere gratitude to my supervisor Henrik Eriksson. You have always guided me with kindness and emotional intelligence which created such a safe space of exploring ideas, good and less good ones. You have permitted me to have own agenda and to be independent, which allowed me to find my identity as a researcher. You have always been a mentor and encouraged to explore possibilities that come with this education. You have definitely contributed to my entrepreneurial pursuits.

My co-supervisor Lex van Velsen, I am greatly thankful for you joining the supervision team and enriching my research with your insights. Your guidance has been essential in my forming as a professional in eHealth. I hope to collaborate with you in the future.

My co-supervisor Andreas Hellström, I would like to thank you for sharing your fresh ideas and helping me ask better questions.

My examiner Ida Gremyr, you have always taken an extra mile in supporting me in crucial moments of my PhD. Not only valuable insights, but also your friendly attitude shall keep guiding me as a great leadership example. Thank you.

Patrik Alexandersson, I have spent my most fruitful and adventurous years of PhD working hand in hand with you. It has been a pleasure and honor to have you as a colleague and friend.

Other dear colleagues, Pernilla Gluch, Dan Andersson, Violeta Roso, Árni Halldorsson, and all the others at SML division, I am so thankful for your support, conversations, beautifully organized events, and a great time together.

My amazing PhD friends, especially Elin Siira, Andrea Birch-Jensen, Sara Dahlin, and Jingshi Min Eriksson, thank you for the care and beautiful memories we created together throughout our journey.

My loving family, Madhu and Tèjas, you both have been the true chargers of my inspiration and strong will throughout these years. Your support, encouragement, and adjustments when accommodating my research needs made this journey peaceful and rewarding.

Finally, I recommend everyone to take the challenge of pursuing a PhD degree!

Sincerely,

Monika Nair

Table of Contents

1 INTRODUCTION 1
1.1 *Background* 1
1.2 *Problem, purpose, and research questions* 1
1.3 *Structure of the thesis* 6

2 FRAME OF REFERENCE 7
2.1 *eHealth* 7
2.2 *Considerations on value of eHealth* 9
2.3 *Evaluation* 10
2.3.1 Evaluation of eHealth 10
2.4 *Standards* 12
2.4.1 Standards in eHealth evaluation 12
2.5 *Translation* 14
2.6 *Interorganizational cooperation* 15
2.7 *Conceptual model* 16

3 METHODOLOGY 19
3.1 *Research strategy* 19
3.1.1 Abductive research approach 19
3.1.2 Process research perspective 19
3.2 *Research design* 20
3.2.1 Single case study 20
3.2.2 Empirical setting 20
3.3 *Research process* 21
3.4 *Data collection and analysis* 22
3.4.1 Mixed methods 22
3.4.2 Study I 23
3.4.3 Study II 25
3.4.4 Study III 26
3.4.5 Study IV 28
3.5 *Reflections on my background and role in DECI and its influence on research* 29
3.6 *Research quality and ethics* 30

4 SUMMARY OF APPENDED PAPERS 37
4.1 *Paper 1: 'Planning a holistic summative eHealth evaluation in an interdisciplinary and multi-national setting: A case study and propositions for guideline development'* 37
4.2 *Paper 2: 'Standards as applied in reality: a case study on the translation of standards in eHealth evaluation practice'* 37
4.3 *Paper 3: 'An Italian business case for an eHealth platform to provide remote monitoring and coaching services for elderly with mild cognitive impairment and mild dementia'* 38
4.4 *Paper 4: 'Identifying the Value of an eHealth Intervention Aimed at Cognitive Impairments: Observational Study in Different Contexts and Service Models'* 39
4.5 *Paper 5: 'Exploring the Use of Evidence From the Development and Evaluation of an Electronic Health (eHealth) Trial: Case Study'* 40
4.6 *Contributions to improving quality in eHealth evaluations* 41

5 DISCUSSION 45
5.1 *RQ1 – How can application of eHealth evaluation standards be improved?* 45
5.2 *RQ2 – How can value of an eHealth solution be conceptualized and measured?* 46
5.3 *RQ3 – How is evidence from an eHealth trial used to support improvements in healthcare?* 47

5.4 ***Theoretical contributions*** **49**
5.5 ***Practical contributions*** **50**
5.6 ***Reflections on actions to improve quality in eHealth evaluations*** **50**
6 CONCLUSIONS AND FUTURE RESEARCH **55**
6.1 ***Conclusions*** **55**
6.2 ***Limitations*** **55**
6.3 ***Future research*** **57**
REFERENCES **59**

List of Figures

Figure 1 – Process framework of the development of cooperative interorganizational relationships (Ring and Van de Ven, 1994)
Figure 2 – Conceptual model for analysis and discussion
Figure 3 – Organizations involved in the project "Digital Environment for Cognitive Inclusion" (DECI)
Figure 4 – Research timeline
Figure 5 – Structure of the value assessment of an eHealth intervention
Figure 6 – PDSA cycle
Figure 7 – A model for improving quality in eHealth evaluations

List of Tables

Table 1 – Methods of data collection and analysis per study and paper
Table 2 – Actions to increase credibility of the research
Table 3 – Actions to increase dependability of the research
Table 4 – Actions to increase validity of the research
Table 5 – Actions to increase reliability of the research
Table 6 – The extended view on eHealth evaluation quality

Annex I:

Table I – Coding of data
Table II – Categorizing (aggregating) of the codes
Table III – Analysis of the intensity of the categories
Table IV – An extract from a comparison between the Health Information Technology Evaluation Toolkit (AHRQ) and Guideline for Good Evaluation practice in Health Informatics (GEP-HI) and the Digital Environment for Cognitive Inclusion (DECI) evaluation process

1 INTRODUCTION

1.1 *Background*

The growth of the Internet has transformed many sectors of economy and industries. Since 2000, healthcare worldwide has harnessed the potential of the Internet and digital technologies (Eysenbach, 2001). Developments in information technology have created opportunities and scenarios to transform healthcare and tackle its three main aims: (1) to improve the health outcomes of the populations, (2) to improve the individual experience of care, and (3) to reduce the per capita cost of care for populations (Berwick et al., 2008).

Among the streams of digitalization in healthcare is eHealth. It represents various types of information and communication technology (ICT) that are employed in healthcare provisions and is considered an effective solution for improving healthcare services (Swinkels et al., 2018). Examples of eHealth applications include wearable technologies, Internet of Things, remote monitoring, virtual care, big data analytics, blockchain, platforms, tools supporting data capture, exchange, and storage, mobile applications (also referred as mHealth), clinical decision support systems, electronic medical records, and telemedicine (World Health Organization, 2018). The applications of eHealth range across the continuum of care, and target diagnosis, treatment, therapies, self-management, clinical decision support, and healthcare management and delivery (Mathews et al., 2019).

Many countries have directed their healthcare policies towards adopting eHealth (World Health Organization, 2016), and investment in the sector is growing exponentially, from both public funds and private investors. For example, app stores contain thousands of health applications, with hundreds more being added daily (Mathews et al., 2019). The increasing volume of eHealth solutions, with many alternatives, makes navigation among them burdensome for interested parties such as patients, caregivers, healthcare organizations, insurers, and regulators. Within this tangled space, it can be challenging to identify which eHealth solutions can provide real value (Mathews et al., 2019).

1.2 *Problem, purpose, and research questions*

The World Health Organization and International Medical Informatics Association emphasized the importance of *summative evaluation* (Lilford et al., 2009), conducted at the end of eHealth trial or implementation, and which purpose is assessing the worth and outcomes of eHealth to the users and system, developing knowledge, and generating evidence for decision-making on eHealth implementations (IMIA Yearbook of Medical Informatics, 2013; Consensus statement of the WHO Global eHealth Evaluation Meeting, 2016). Moreover, in supporting decision-

making by people responsible for eHealth implementation and policies, evaluation quality is a key in order to utilize the results produced by evaluation studies (Mookherji et al., 2015; Dick et al., 2020). Quality in eHealth evaluation is achieved when a study captures contextual specifics (Andargoli et al., 2017), multi-stakeholder perspectives are considered (Greenhalgh and Russel, 2010), research design is adequate (Pham et al., 2016), methodology is based on scientific methods and standardized approaches (Dick et al., 2020), and data are handled ethically (Mechael et al., 2019).

However, despite the vast efforts of the scientific community in supporting evaluators with evaluation methodologies, a number of concerns are prominent in the field. First, there is a growing concern of 'pilotism', a phenomenon that refers to effort and resources being wasted when the outcomes of a pilot study are not utilized (Tomlinson et al., 2013; Andreassen et al., 2015; Urueña et al., 2016). Second, evaluations can be affected by the culture, assumptions, values, and agendas of evaluators (Stufflebeam, 2001; Chouinard and Cousins, 2009). Third, evaluation studies can be burdened by the involvement of several parties; for instance, aligning goals and ideas of multiple parties can be a challenge (Greenhalgh and Russell, 2010; Vangen and Huxman, 2011). Fourth, insufficient quality of evaluation studies has been reported as one factor that contributes to pilotism (Mookherji et al., 2015), referring to an overly narrow scope of evaluations and their scientific rigor, which to a large extent are determined during evaluation planning. Therefore, there is a need to better understand the problems surrounding quality of evaluation studies in order to increase credibility in the evaluation outcomes and their use (Mookherji et al., 2015).

Researchers have discussed a number of challenges related to quality in summative eHealth evaluations. Several scholars have noted the lack of generalizable knowledge produced by eHealth evaluations (Dick et al., 2020; Kip et al., 2021). Among the challenges that give rise to such lack are insufficiently rigorous research methodologies or study designs (Dick et al., 2020). Further, limited by financial resources, evaluators often pursue an insufficient scope of evaluation themes for informed decision-making, thus reducing the possibility for an eHealth solution to be adopted in real practice (Andargoli et al., 2017). Another question is whether evaluators should be external or internal to the organization. External roles might create conditions for an un-biased assessment (Mookherji et al., 2015), while being internal and actively engaging in the intervention and evaluation activities might help obtain deeper insights into the studied phenomenon (Greenhalgh and Russel, 2009).

Another challenge is the lack of a single best method to perform an eHealth evaluation (Dick et al., 2020). The research methods and outcome measures are sometimes tailor-made and context-specific, making it impossible to have common denominators within related studies (Glasgow, 2007; Dick et al., 2020). Application of evaluation standards is seen as a pathway to increase scientific rigor and quality in evaluations (Mookherji et al., 2015; Cowie et al., 2016; Dick et al., 2020). A need to create more and better standards has been reported in practitioner surveys (Mookherji et al., 2015; Dick et al., 2020). However, there is a growing concern that

the actual usage of the evaluation standards is insufficient (Mookherji et al., 2015; Cowie et al., 2016; Dick et al., 2020).

Furthermore, due to lengthy review processes by the scientific outlets, the relevance of published evaluation studies can be reduced, and eHealth technology can become outdated (Whittaker et al., 2012). In addition, it is not possible to attribute benefits to changes in some outcomes (such as behavior) to eHealth alone, due to its embeddedness in a wider context with many confounding factors (Mechael et al., 2019). Also, changes in outcomes are not stable and they change over time (Greenhalgh and Russel, 2009). The true benefits of an eHealth intervention might not manifest through the pre-planned outcome measures and might actually be something else (Greenhalgh and Russel, 2009).

Another concern in the eHealth evaluation research is the appropriateness of the methodological approach (Greenhalgh and Russel, 2010; Dick et al., 2020). The choice of methodology depends primarily on the ontological assumption that a researcher follows, either explicitly or implicitly (Greenhalgh and Russel, 2010; Nykänen et al., 2011; McNair, 2016). When an eHealth evaluation is conducted under the *positivist* approach (Orlikowski and Baroudi, 1991), objectivity in the assessment of "reality" is assumed. This approach resembles the Health Technology Assessment (HTA) (Kazanjian and Green, 2002), which is supported by a formal framework and summative by nature (Nykänen et al., 2011). Studies that employ this approach often follow a randomized controlled trial (RCT) design and this design composes a substantial part in eHealth evaluations (e.g., 80 percent of mHealth evaluation studies (Pham et al., 2016)). However, scholars have criticized the application of a positivist approach to eHealth evaluation research (Greenhalgh and Russel, 2010; Robertson et al., 2010; Mechael et al., 2019; Dick et al., 2020). They argue that such an approach cannot capture the dynamic and socio-technical nature of eHealth and the surrounding context. Moreover, they noted the wrong assumption formed in employing this approach; that is, research cannot be conducted in a controlled environment, supposing that eHealth is embedded in a social context that needs to be accounted for during the evaluation (Andargoli et al., 2017; Dick et al., 2020). Greenhalgh and Russel (2010) proposed alternative eHealth evaluation methodologies based on an *interpretivist* approach (Klein and Myers, 1999). The basis for these propositions is that eHealth evaluation research could be considered as a "social practice rather than as a scientific testing" (Greenhalgh and Russel, 2010).

Currently, the dominant advise for eHealth evaluation research methodology is to apply "methodological pluralism" (or mixed-method approach) (Lilford et al., 2009; Andargoli et al., 2017; Dick et a., 2020). This approach is considered a solution to the criticism that quantitative research methods do not capture the complexity of the socio-technical environment of eHealth implementation. Hence, the mixed-method approach, which combines positivist and interpretivist approaches, is preferred in eHealth research because it provides the possibility to ask the "why" questions in addition to the traditional "what," "where," and "who" questions (Andargoli et al., 2017). The consideration of which methodological approach to follow is an

important aspect in planning an eHealth evaluation because when the methodological approaches differ between related studies, the comparability and generalizability of such knowledge are limited.

Given the challenges discussed above, the ***purpose*** of this thesis is to study how summative eHealth evaluations can be improved to support the determination of eHealth value in a specific context, as well as the use of evidence produced during eHealth interventions.

Following this purpose, the thesis is guided by three research questions. First, the thesis addresses improvement of application of standards used in summative eHealth evaluations, as they are the "tools" of evaluation (Research Question 1). The thesis then argues for the need for a unified approach (the unified 'what') towards evaluating eHealth and proposes approaching it through a concept of value and suggests a framework for evaluating it (Research Question 2). Finally, the thesis analyzes the 'afterlife' of evidence created by- and through a process of eHealth intervention and its evaluation (Research Question 3) to understand usefulness of the evidence leading to improvement suggestions for future eHealth evaluations.

Research Question 1

Standardization of summative eHealth evaluations is considered one of the possible pathways to increase quality (Proudfoot et al., 2011; Mookherji et al., 2015; Cowie et al., 2016; Dick et al., 2020). Usage of standards, such as guidelines, evaluation frameworks, and standardized metrics creates trust in the methodology and findings and leads to various degrees of methodological uniformity between different studies (useful in cross-country evaluations and systematic reviews) and enhances generalizability (Lilford et al., 2009; Mookherji et al., 2015; Cowie et al., 2016).

Although the use of standards has been promoted by different scholars and organizations, it seems that standards are not always applied in empirical evaluation studies (Janssen et al., 2013a; Mookherji et al., 2015). Previous research has focused mostly on creation or improvement of different standards, but not on assessing their actual application or extent of their feasibility and applicability in a particular context. Studying the application of standards in practice of eHealth evaluations can reveal the deficiencies that create conditions for problems with evaluation quality. Since the standards used in eHealth evaluations can originate from medical, information technology, or business fields, it is crucial to understand how applicable (being useful for generating meaningful evidence) the standards are in order to achieve quality in eHealth evaluations. This has been attempted for other standardized improvement concepts, such as Lean in healthcare, that have been questioned in terms of feasibility, if not translated to fit a particular context (e.g., Wæraas and Sataøen, 2014; Andersen et al., 2014). The importance of evaluating the effectiveness of standards, comparing, and improving existing ones was emphasized by professional medical societies (Cowie et al., 2016).

RQ1: How can application of eHealth evaluation standards be improved?

Research Question 2:

In healthcare, the concept of value has risen to prominence partly due to the rise of an idea of value-based healthcare (Porter and Teisberg, 2006), in which value is defined as the ratio between health outcomes and cost. On a more general level, value is well established in the field of service management (Ramirez, 1999; Grönroos, 2008; Gummerus, 2013). However, the concept of value is vague when it comes to the context of eHealth. Appropriate methods and evaluation criteria to assess value of eHealth are also missing (WHO, 2016). In talking about value of an eHealth solution, some evaluators have focused on clinical efficacy or behavior change of patients or professionals. Some have singled out or added to the equation economic outcomes (such as cost-effectiveness), building on the dominant approach of Health Technology Assessment (HTA), an evaluation framework primarily created for evaluations of pharmaceutic interventions (Bergmo, 2015). Others take a completely qualitative approach and evaluate perceptions of the value perceived by various stakeholders (Runz-Jørgensen et al., 2017). This variety of value conceptualizations, and ways of operationalizing it, might delimit the understanding of value and might overlook outcomes beyond the ones delivered by traditional HTA or qualitative approaches.

RQ2: How can value of eHealth solution be conceptualized and measured?

Research Question 3:

Similar to evidence-based practice in medicine, the implementation and evaluation of eHealth solutions should produce credible evidence for decision making (Ammenwerth and Rigby, 2016; Rigby et al., 2018). The evidence concerned includes research evidence, professional experience, and patient preferences (Greenhalgh et al., 2014; Djulbegovic and Guyatt, 2017). A summative evaluation of eHealth interventions, a highly resource consuming activity, produces a considerable share of evidence. However, there is a growing concern that evidence is poorly used when making clinical and policy decisions in the context of eHealth (Cohen et al., 2015; Koppel, 2018; Alla et al., 2018). Hence, there is a need to assess how evidence from eHealth implementations is used by different stakeholders and why they find the evidence use problematic.

RQ3: How is evidence from an eHealth implementation used to support improvements in healthcare?

1.3 *Structure of the thesis*

This thesis is based on five research papers. Chapter 2 includes a description of a frame of reference, including concepts like eHealth, evaluation, value, standards, and relevant theories used in the studies. Chapter 3 outlines the research process, context, and methodology, followed by an explanation of the research methods used in each study, and a discussion of research quality issues. Chapter 4 provides brief summaries of all the papers included in this thesis. A discussion of theoretical and practical research implications follows in Chapter 5, and the thesis ends with conclusions, limitations, and suggestions for future research described in Chapter 6.

2 FRAME OF REFERENCE

This chapter introduces key theoretical concepts such as eHealth, evaluation science and evaluation of eHealth to provide a context for the purpose of the thesis. Theory on evidence use is introduced to provide a theoretical framework when studying how the evidence from eHealth trials are used and to shine light on the problem of pilotism. Then, various considerations on value of eHealth are elaborated to problematize the ambiguity of the value concept and its usage in this field. After that, theories on standardization, translation, and interorganizational cooperation are introduced in order to inform the study on standards' applicability in the context of eHealth.

2.1 *eHealth*

"eHealth" has become a prominent term since 1999 and has been marketed as harnessing the opportunities of the internet and e-commerce within healthcare (Eysenbach, 2001). However, scholars have recognized early that eHealth is not limited to technology and its development (Eysenbach, 2001), and that its boundaries are not delimited (Shaw et al., 2017). Among the most accepted (Pagliari et al., 2005) and earliest definitions of eHealth is Eysenbach's (2001, p. 1):

"e-health is an emerging field in the intersection of medical informatics, public health and business, referring to health services and information delivered or enhanced through the Internet and related technologies. In a broader sense, the term characterizes not only a technical development, but also a state-of-mind, a way of thinking, an attitude, and a commitment for networked, global thinking, to improve health care locally, regionally, and worldwide by using information and communication technology."

In this thesis, I followed the view on eHealth suggested by Kaplan and Shaw (2004) that draws more towards the process and complexities of implementing eHealth. Kaplan and Shaw stated that eHealth is a complex innovation implemented in a socio-technical environment (Kaplan and Shaw, 2004), and implementations of eHealth may demand that processes be reorganized, professionals re-trained, and individual habits adjusted. This view allows for understanding that eHealth solutions do not operate in isolation from social settings that are dynamic and particular depending on the context.

According to Shaw et al. (2017), the dominant types of eHealth are: *(1) Health in your hands*: the use of eHealth technologies to monitor, track, and inform health; *(2) Interacting for health*: the use of technologies to communicate between stakeholders in health; *(3) Data enabling health*: the collection, management, and use of health data sources. The *health in your hands* domain refers to consumers' own observation of their health data (for example, through different web and mobile applications) or enabling convenient access to health-related

information (Shaw et al., 2017), thus improving health education (Khoja et al., 2013). The *interacting for health* domain concerns ICT-enabled interactions between professionals, or in combination with patients. Based on tele- or video-conferencing, SMS or push notifications, various communication platforms (such as telehealth, fitness coaching and professional support, integrated care, and social media) can be assigned to this domain. The *data enabling health* domain refers to the collection, analysis (also powered by artificial intelligence and machine learning), and use of big data (such as diagnostics, and predictive and precision healthcare). The analysis can concern different scopes, ranging from an individual level to the entire population (Shaw et al., 2017), thus enhancing research potential and improving care and its safety (Khoja et al., 2013). All of these eHealth developments are the sources of business opportunities that can be delivered through innovative business models (Shaw et al., 2017).

Due to the investments that governments and private investors have poured into the fields of eHealth and life science innovation in general, the number of available digital health solutions continues to grow (Mathews et al., 2019). In this light, responsible institutions worldwide (such as the European Commission and the Food and Drug Administration in the US) continue to tighten regulations and put control measures on the solutions entering the markets to ensure quality and safety (Regulation (EU) 2017/745, 2017). Still, the volume of such solutions continues to increase, many of them offering overlapping and competing benefits from a technology, health condition or lifestyle, or population perspectives. Such a busy landscape creates difficulties for adopters in navigating and deciding which solutions provide true or best value (Mathews et al., 2019). In addition, a number of factors slow down the adoption of eHealth in the healthcare sector (Barlow, 2016). First, the healthcare sector is strictly regulated. Second, eHealth innovations are embedded in organizational and cultural context and require changes in processes, skills, attitudes, or behavior (Ariens et al., 2017). Third, implementations of an innovation in healthcare, similarly to other industries, can be multi-purpose, aiming to increase the quality of care in a safe way and at a lower cost than traditional care (Barlow, 2016). Fourth, there can be a multiplicity of users of the innovation – healthcare professionals, patients and their caregivers – and all these actors need to be convinced of the innovation in order to get it adopted (Tidd and Bessant, 2014). Fifth, the payer for the healthcare service is usually not its user, but a government or an insurance company. Sixth, the main beneficiary of a service is usually not the payer. For example, a municipality pays for the service, but healthcare benefits go the patient and cost savings go to the insurer (in some countries). Seventh, manufacturers of the technology innovation are increasingly important parties in healthcare service delivery, and quality of healthcare service largely depend on these organizations. Eighth, there is a big gap between research and development in healthcare and its translation to practice, which leads to substantial waste of effort and resources (Lomas, 2007; Nicolini et al., 2008).

2.2 *Considerations on value of eHealth*

A discussion on what value of eHealth is and how to determine it is important in order to have evaluation approaches built on unified interpretations of value. The term 'value' is frequently referred to when discussing worth of an eHealth solution. These discussions can appear in a number of situations. For example, value is assessed when a regulatory authority considers granting an approval enabling a manufacturer to sell the solution in particular markets. In determining which eHealth solutions are worthy to be commercially available in the markets, regulatory authorities base their decisions on clinical efficacy, usability, and safety (European Commission, MEDDEV 2.7/1, 2016). Another example is when an organization assesses which eHealth solution to adopt. Besides looking for the best health outcomes for the money spent, potential adopters try to understand how disruptive the technology is to their workflows (Barlow, 2016). eHealth researchers have also used the concept of value quite loosely, which has resulted in a range of possible specifications of value in this context. For instance, scholars refer to cost-effectiveness or cost-benefit results (Bergmo, 2018), satisfaction by patients or healthcare staff of using a solution (Runz-Jørgensen et al., 2017). These examples show that the usage of the value concept in the context of eHealth is flexible and the concept itself is vague.

In a broader context, value-based healthcare is a concept that is increasingly diffused in healthcare systems (Young, 2015; Elf et al., 2017). It is conceptualized as patient outcomes compared to the costs of care (Porter and Teisberg, 2006). The existence of a producer and perceiver is embedded in the concept, meaning that there should be a producer and a perceiver of value in order for it to exist (Gummerus, 2013). Furthermore, value involves reciprocity; it is created together with a customer and the service provider can facilitate the value creation by inviting the customer to co-create it (Grönroos, 2008).

Various stakeholders can take part in value creation activities, such as patients, organizations, insurers, policy makers, and entrepreneurs. The expected value from eHealth solutions and their implementations is different for these actors. Patients often expect health benefits or easier or more accurate management of their disease; healthcare professionals expect to have better tools to provide care and advantages in reputation or image of their organization; insurers see value through cost-effectiveness perspective; policy makers want to see a holistic view on the outcomes of the implementations; and entrepreneurs are interested in revenues and scientific validation from implementations of their solutions (a positive business case) (Swinkels et al., 2018). Some scholars also refer to co-production in healthcare (Batalden et al., 2016). For example, patients are engaged in self-management of health and report the data to the healthcare organization that makes decisions. Furthermore, perceptions of value have been noticed to be dependent on the context (Gummerus, 2013). This means that value is perceived through an experience of a particular person in a particular situation and in a certain contextual setting. This relativity and vagueness of the value concept create complexity in assessing the worth of eHealth solutions.

2.3 *Evaluation*

Evaluation is an important area of science and practice, referred to as evaluation science. This area is usually guided by four main goals: (1) to improve something (such as practice, a solution, policy), (2) to monitor compliance, (3) to assess merit and worth, and (4) to develop knowledge (Mark et al., 2000). The emergence of the field was propelled by the increasing demands for accountability in democratic societies (Donaldson and Lipsey, 2006). An evaluation is expected to guide decision-makers toward making better decisions regarding service or policy, and to improve quality of services and investment decisions (Mark et al., 2006).

Theories, frameworks, methods, and tools comprise an important base of evaluation science (Mark, 2003; Donaldson and Christie, 2006). The role and utility of theory and theoretical approaches in evaluation practice is a debatable topic. Some scholars have argued that a theory-based approach to evaluation is essential, as it guides evaluation practice and provides knowledge base and common denominators for evaluation theorists and practitioners (Shadish, 1998; Mark, 2003). By contrast, those opposing the perspective of relying on theory in evaluation argue that it is almost always impossible to perform an evaluation properly according to a theory, and failed interpretations can lead to counter-productive outcomes (Stufflebeam, 2001). Also, the social reality is so complex that it is naïve to assume that one can pre-determine aspects to be assessed (Stufflebeam, 2001). Moreover, aiming for a theory-based evaluation is, at times, an impractical idea; that is, trying to evaluate every item that can be derived from a theory takes time and resources (Scriven, 1998).

The key theories involved in evaluation science are evaluation, program, and social science theories (Donaldson and Lipsey, 2006). Evaluation theory is highly prescriptive and builds on principles and standards that provide methods, tools, and guiding frameworks to evaluation practice (Alkin, 2004). Social science theory is instrumental in understanding the way individuals function and behave. It helps in evaluation design and offers a context for interpreting evaluation results (Donaldson and Lipsey, 2006). Program theory helps create a model of how a program, intervention, or treatment should work. Particular elements of change are assumed to affect results through certain processes and under certain conditions (Bickman, 1987; Lipsey, 1993). For the purpose of knowledge development while interpreting evaluation results, program and social theories are more significant, whereas evaluation theory is largely instrumental in practice.

2.3.1 Evaluation of eHealth

Ammenwerth et al. (2004) defined an evaluation of eHealth as "the act of measuring or exploring properties of a health information system (in planning, in development, in implementation, or in operation), the result of which informs a decision to be made concerning that system in a specific context." This notion resembles evidence-based medicine (Evidence-

based Working Group, 1992), where decision making should "rely on an explicit evidence derived from rigorous studies on what makes systems clinically acceptable, safe and effective – not on basic science or experts alone" (Wyatt, 2016, p.15).

The main types of eHealth evaluation are *formative* and *summative* evaluations (Lincoln and Guba, 1986). The aim of a *formative* evaluation is to provide feedback (such as staff response) during the implementation or design of an eHealth solution. Such evaluation is iterative, producing immediate but less generalizable knowledge. *Summative* evaluation, which is the main focus of the present thesis, is performed at the end of the implementation; its purpose is to assess the outcomes of eHealth to the users and system, and to provide a somewhat generalizable knowledge that is useful in decision making (Lilford et al., 2009). Hence, evidence should have utility in decision-making, meaning that it should be used. When the expectation is to use evidence in the decision-making, it refers to the instrumental use, which is the direct use of the information in decision-making and taking action, in order to change the existing practice (Nutley et al., 2007; Alkin and King, 2017; Weiss, 1999; Leviton and Hughes, 2016). Conceptual use refers to a non-direct use of information and perspectives to enhance understanding. Strategic or symbolic use occurs when the evidence is brought up to support or confront an existing idea or decision (Nutley et al., 2007; Alkin and King, 2017; Weiss, 1999; Leviton and Hughes, 1981). The use types of evidence will be applied when exploring how the evidence from an eHealth trial is used in supporting improvements in healthcare (Research Question 3).

Several aspects of a summative evaluation could be highlighted in the context of eHealth: (1) Summative eHealth evaluations are performed after all the research data have been collected, and the situation is 'frozen in the moment' (Lilford et al., 2009). In other words, a summative evaluation is a snapshot of reality that is otherwise highly dynamic (Barlow, 2016). (2) In that single snapshot, previous research recommends viewing the impact to stakeholders, processes, and economics from various angles (evaluation domains), as recommended in the evaluation frameworks (e.g., Lampe et al., 2009; Kidholm et al., 2012). (3) Although a summative eHealth evaluation does not aim to reflect upon the process of value creation, it has been acknowledged that a multiplicity of actors or stakeholders are related to the value creation, and their perspectives need to be captured simultaneously (this reflects in evaluation frameworks, e.g., Andargoli et al., 2017). (4) Moreover, an eHealth solution is embedded in clinical care that can be viewed as service connecting two ends: a consumer (a patient or a caregiver) and a provider (a healthcare organization) (Grönroos, 2008). Complexity also increases because of a manufacturer of an eHealth solution, which can be a complementary service provider in addition to clinicians (for example, a technology company provides software and creates content to patients that is prescribed by clinicians).

2.4 *Standards*

Research on how the application of eHealth evaluation standards can be improved (Research Question 1) can benefit from considering the conceptual understanding of what a standard is. Allen and Sriram (2000) identified three categories of standards: de facto, regulatory, and consensus. The category depends on a standard's origin and creation processes. De facto standards are those that are widely adopted but not regulated (for example, a PC keyboard that is defined by the first six characters on its upper left side: QWERTY). Regulatory standards are issued by regulatory institutions with a goal of creating uniformity in particular processes of an industry (for example, standards that regulate safety requirements for particular workplaces or occupations). Consensus standards are issued by local or international bodies to encourage users to voluntarily conform with a standard (such as standards issued by the International Organization for Standardization [ISO]).

An expected outcome of developing a standard is to make users (organizations or individuals) aim for a particular result or process (Brunsson and Jacobsson, 2000). However, the issuing body of a standard (or a standardizer) has no formal authority or sanctioning power over the adopters, leaving the adherence to a standard dependent on the free will of users (for example, ISO standards). What standardizers offer in a standard is only a recipe and guidance to the adopters. A common association to the function of a standard is the creation of similarity and uniformity (Brunsson and Jacobsson, 2000), and compatibility (Farrell and Saloner, 1995). This idea comes from the assumption that users adhering repeatedly to a standard creates similarity over time. Correspondingly, once many users adopt a standard, similarity across space is generated (Brunsson and Jacobsson, 2000). Farrell and Saloner (1985) suggested that the value of standardization comes with the economies of scale. There is also a negative side of standardization: once a standard becomes widely diffused, it can fixate the practice, which becomes a barrier for accepting better practices (Farrell and Saloner, 1985; Brunsson and Jacobsson, 2000). It is especially problematic if the standard becomes obsolete and needs to be revised. Similarly, Farrell and Saloner (1985) argued that standardization can hinder innovation because of users' potential unwillingness to switch to a new standard.

2.4.1 Standards in eHealth evaluation

Because eHealth is a crossroad between the medical, social, and information systems fields, a considerable number of available standardized approaches (that is, standards) can be applied to eHealth evaluation. The present thesis deals with the following types of standards: (1) eHealth evaluation planning guidelines offering guidance in planning an evaluation study; (2) eHealth evaluation frameworks offering a structure in terms of evaluation themes; and (3) standard outcome indicators, such as standard scales and questionnaires, designed to measure individual outcomes of a studied intervention. These types of standards are considered *consensus* (voluntary) standards (Allen and Sriram, 2000).

(1) eHealth evaluation planning guidelines

To guide evaluators in the initiation of an eHealth evaluation process, guidelines for evaluation planning have been published. Examples of these guidelines are the Health Information Technology Evaluation Toolkit, or AHRQ toolkit (Cusack et al., 2009); Design and Evaluation guidelines for mental health technologies by Doherty et al. (2010); Guideline for Good Evaluation practice in Health Informatics, or GEP-HI (Nykänen et al., 2011); Model for Assessment of Telemedicine Applications, or MAST model (Kidholm et al., 2012); European Commission's MEDDEV 2.7/1 (European Commission, 2016). The AHRQ toolkit offers a step-by-step guidance to evaluation planning, especially in the operationalization of methods. The MAST model, GEP-HI, and Doherty et al.'s (2010) unnamed guideline provide a list of elements to consider while planning an eHealth evaluation.

(2) eHealth evaluation frameworks

For the study design, evaluation needs to have a structure of evaluation topics and can concern perspectives of different stakeholder groups, such as patients, caregivers, healthcare professionals, and policymakers (Evans, 2003; Nykänen et al., 2011). To address the needs of decision-makers in making informed decisions and improve uniformity among eHealth evaluation studies (Ammenwerth, 2004; Proudfoot et al., 2011; Janssen et al., 2013a), different evaluation frameworks have been developed by scholars, who agree on the imperativeness of unifying the approach to evaluation (Kaplan and Shaw, 2004). The variety of frameworks is created when the scholars suggested eHealth type-specific frameworks, such as the framework for telemedicine (e.g., MAST [Kidholm et al., 2012]) or telecare (Williams and Doughty, 2007). These frameworks are differentiated by their methodological approach, namely, formative (e.g., Performance of Routine Information System Management, or PRISM [Aqil et al., 2009]), summative (e.g., MAST [Kidholm et al., 2012]), or mixed (formative and summative) frameworks (e.g., van Gemert-Pijnen et al., 2011). However, systematic literature reviews (Yusof et al., 2008; van Gemert-Pijnen et al., 2011; Andargoli et al., 2017) concluded that evaluation frameworks are insufficient and do not address all relevant elements (for instance, they miss to address the questions of what, how, when, why and who of the evaluation, or neglect the role of context (Andargoli et al., 2017)). Moreover, critics of the standard evaluation frameworks have argued that the usefulness of such frameworks is limited because of the different contexts on which each framework was based (Bates and Wright, 2009) and because no framework can suit all eHealth evaluation studies (Kaplan and Shaw, 2004). However, new and improved eHealth evaluation frameworks continue to be published to support a unifying approach (for example, Greenhalgh et al.'s [2017] Non-adoption, Abandonment, Scale-up, Spread, and Sustainability framework [NASS]).

(3) Standardized metrics

The eHealth evaluation research themes and methodology have been operationalized through

outcome indicators (Nykänen et al., 2011). Numerous standard questionnaires and scales have been developed to support measurement of the evaluation themes. Some metrics particularly target eHealth (for example, the eHealth literacy scale [eHEALS; Norman and Skinner, 2006] or eHealth impact questionnaire [Kelly et al., 2015]). Others are generic (such as the Patient Satisfaction Questionnaire [PSQ-18; Marshall and Hays, 1994]). Still others are various medical scales that evaluate the impact of an eHealth solution on medical outcomes. The validity of such standards of measurement of outcome indicators is essential as it increases credibility of the measurement tools that support decision-making regarding the adoption of an eHealth solution. Therefore, different scholars have sought to validate these standards (examples include validations of eHEALS in Dutch [Van Der Vaart et al., 2011] and Italian [Diviani, 2014] contexts).

When different standards are considered for inclusion in the methodology of an eHealth evaluation study, their fit to the study is examined. These activities determine whether a standard will be included in the methodology, how it will be used, and what adaptations to a standard might be needed. Section 2.5 describes these processes.

2.5 *Translation*

When one or several stakeholders work with an object or an idea, they may "tailor the object in such a way that it caters for these people's explicit interests" (Latour, 1987). Such a process is called translation (Latour, 1987; Røvik, 2007; Czarniawska and Sevón, 2011). Organizational studies have noted that management ideas change when they are deployed to a specific environment (Hellström et al., 2010; Wæraas and Sataøen, 2014). Røvik (2007) suggested that ideas are applied to a particular context using four main strategies of translation: copying, addition, omission, and alteration. *Copying* aims for accuracy regarding the original content of an idea. *Addition* indicates the extension of the original content of an idea to align with the peculiarities of a context. *Omission* means that one or more features of the original content of an idea are removed when it is ill-fit with the context or other conditions (for example, when the value of some features of an idea is questionable in a particular context or due to financial constraints). *Alteration* means that the original content of an idea is largely changed. Alteration can be understood as a strategy that is contrary to copying. The level of detail in an idea or model to be translated can influence the strategy selected (Røvik, 2007). The selection of a strategy of translation can also depend on the characteristics of a field or industry (Røvik, 2007; Wæraas and Sataøen, 2014). These translation strategies shall be used when interpreting how the evaluation standards were selected and adjusted to fit the circumstances of the research project (Research Question 1).

At times, translation processes do not depend on a single actor, and decision-making on the outcomes of translation can be a multiparty activity. Section 2.6 introduces the processes that take place during collaborative activities.

2.6 *Interorganizational cooperation*

Growing international funding in eHealth research and implementation (European Commission's Directorate General for Communications Networks, 2016) has led to studies becoming increasingly multidisciplinary and international (Greenhalgh and Russel, 2010). This has made it more important to assess the impact of interorganizational cooperation in eHealth evaluation research. The process framework of the development of cooperative interorganizational relationships developed by Ring and Van de Ven (1994) presents key processes that occur when different parties cooperate (see Figure 1). These processes involve negotiation, commitment, execution, and assessment. Cooperation happens in iterations, and the outcomes of cooperation are evaluated for reaching its intended goals. To move forward with the cooperation, diverse ideas of different actors need to be aligned (Greenhalgh and Russell, 2010; Vangen and Huxman, 2011).

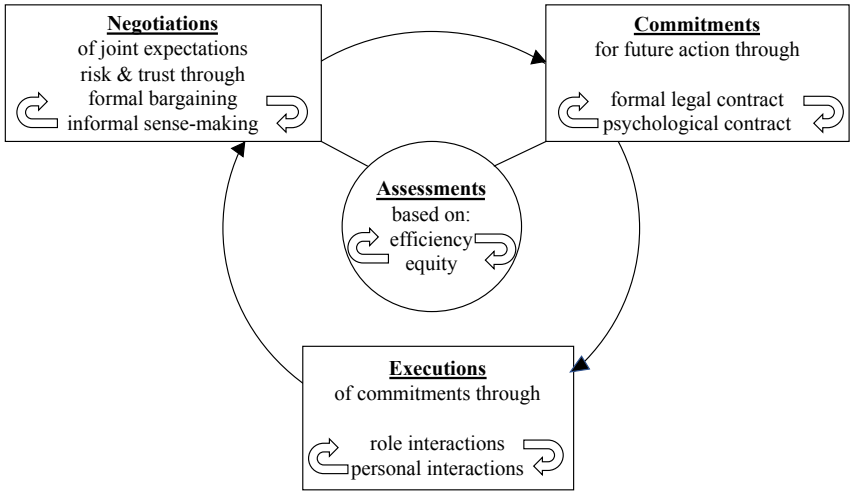

Figure 1. Process framework of the development of cooperative interorganizational relationships (Ring and Van de Ven, 1994)

During the *negotiations* stage in the framework (Ring and Van de Ven, 1994), the cooperating parties aim to align and create joint expectations and to reach a consensus on a subject or terms of cooperation. When the parties arrive at the *commitments* stage, the consensus and the commitment to a future action has been reached. The commitment is manifested in a written document or in a verbal agreement. Once the achieved commitments are realized in action, the parties have reached the *executions* stage. The need for potential updates to cooperation can be *assessed* at any of these stages because, among other things, changes in the contextual elements or changed positions of the parties can lead to renegotiations (Ring and Van de Ven, 1994).

The process framework (Figure 1) is applicable to a research consortium because it is one possible kind of an interorganizational relationship (Ring and Van de Ven, 1994). Moreover, as Greenhalgh et al. (2004) pointed out, the research related to innovation in service organizations, including healthcare, lacks a process view. Such a view can enrich the

understanding of the different elements surrounding the phenomenon and help improve those elements. In this thesis, the framework is employed to examine the translation of eHealth evaluation standards as enacted through the processes of interorganizational cooperation. In the case of evaluations of eHealth interventions, the 'formal legal contract' (commitment phase) could be interpreted as a documented evaluation plan based on the consensus between the stakeholders. The assessment phase might also be less applicable in this context, since some designs of summative evaluations (such as a randomized control trial) are less flexible for changes.

This framework shall be used when interpreting how different stakeholders have arrived at common decisions regarding inclusion of different evaluation standards in the research study (Research Question 1).

2.7 *Conceptual model*

An integration of the above-described theories and concepts makes it possible to explore how to improve quality in summative eHealth evaluations. An additional focus is placed on supporting the identification of eHealth value in a particular context and the use of evidence produced during eHealth interventions. The theories can be summarized in a conceptual model (see Figure 2).

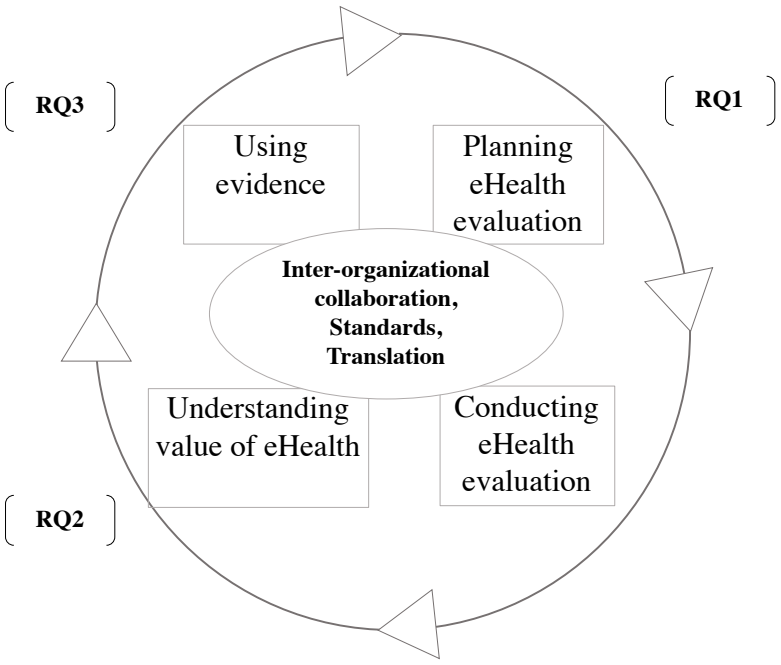

Figure 2. Conceptual model for analysis and discussion

Quality problems in eHealth evaluations can be identified and prevented in all the phases of the evaluation, from planning through to studying how the created evidence is used. As the arrows

in the model depict, it might be possible to revisit every activity for preventing or solving the problem that manifests in the next step, and that might diminish quality of the evaluation. Evaluations are increasingly interdisciplinary, and collaboration between different actors is central, as it can affect evaluation through culture, assumptions, values, and agendas of collaborating parties (Stufflebeam, 2001; Chouinard and Cousins, 2009). Challenges emerging from collaborative activities in a multiparty research have been noted in the previous research, such as alignment of goals and ideas (Greenhalgh and Russell, 2010; Vangen and Huxman, 2011). Therefore, the inter-organizational collaboration is depicted as an important element in the conceptual model, as problems stemming from the collaboration and leading to evaluation quality issues shall be explored to identify strategies to prevent or solve those issues. Using standards during evaluation planning is vital for quality of evaluation (Mookherji et al., 2015; Cowie et al., 2016; Dick et al., 2020) (depicted as 'Standards' in the conceptual model). The translation process and translation strategies selected (Røvik, 2007) determine how a standard will be used, if at all, and whether it will be adhered to in a particular evaluation study (depicted as 'Translation' in the conceptual model). Interpreting the created evidence determines value of an eHealth solution in a particular context. Finally, the created evidence can be used in different ways, such as instrumentally, conceptually, or symbolically (Nutley et al., 2007; Alkin and King, 2017; Weiss, 1999; Leviton and Hughes, 1981), to serve different purposes (depicted as 'Using evidence' in the conceptual model). By studying how evidence is used by the stakeholders, it can be possible to identify the supporting strategies for improving quality in eHealth evaluations.

3 METHODOLOGY

The following sections provide more details on the research strategy and design, the studied case, research process, data collection and analysis in the conducted studies. The chapter ends with some considerations regarding the impact of my own background and role in the research project to the research outputs.

3.1 *Research strategy*

3.1.1 Abductive research approach

Taking an abductive research approach means using inductive and deductive approaches interchangeably. Abduction does not start with pre-defined questions and theoretical lens, nor does it start from a 'blank page', without previous understanding (Langley et al., 2013). Research described in this thesis has been empirically driven, but the empirical data was used to contribute to knowledge through conceptualization (Schwarz and Stensaker, 2014). Following the objectives and activities in the project that served as an empirical setting for this thesis, practical issues manifested. Curiosity has led to explorations of whether those issues had been addressed and theorized in previous research, including what methods could help in analyzing those issues. This helped to formulate research questions and to design the studies. Then, to analyze the empirical data, studies employed existing theoretical frameworks from eHealth or adjacent fields. It means that the empirical data were an instrument for theorizing (Van Maanen et al. 2007). By abductively alternating between empirical data and theoretical frameworks, the data were structured and explained through the theory, new conceptualizations were made, and gaps were identified (Van Maanen et al. 2007).

3.1.2 Process research perspective

Exploration of how application of the eHealth evaluation standards can be improved (Research Question 1) employed a process research perspective (Langley, 1999; Langley and Tsoukas, 2017). Greenhalgh et al. (2004) noted that an understanding of the processes surrounding innovations in healthcare is not only crucial, but also largely missing in the published literature. Such an understanding can help explain how innovations are implemented and adopted by users of an organization. Evaluation is an important part of eHealth research and implementation projects and has implications concerning evidence for decision-making, which can lead to the adoption and sustainability of an eHealth solution. An eHealth evaluation process consists of the phases of preliminary outline, study design, operationalization of methods, project planning, execution of the evaluation study, and completion of the evaluation study (Nykänen et al., 2011). Although there are different approaches to studying a process (Langley and Tsoukas, 2017), the concept of a process used in this thesis referred to sequential activities that lead to an implementation of a particular goal (Saldaña, 2003).

3.2 *Research design*

Building on the evaluation quality problems outlined by the previous research, an in-depth exploration of the barriers and potential solutions to those problems was sought. This could be achieved through a single case research design, which was selected for this thesis. All of the conducted studies were based on a single case study design executed through a multinational and interdisciplinary research project (described below).

3.2.1 Single case study

The research conducted in all of the studies in the current thesis were built on a single-case study design. Such a design allows a multi-faceted in-depth study and assessment of a phenomenon in a unique setting (Yin, 2014; Flick, 2014). The case of this thesis concerned a single research project with nine collaborating stakeholders, and four clinical trial sites in different countries (the project is described in Chapter 3.2.2 "Empirical setting"). The case was focused on eHealth evaluation, which was studied longitudinally, from planning an evaluation to using its results. A longitudinal view of a single case was chosen (1) to take a deep dive into the particularities of stakeholders' collaboration and its impact on eHealth evaluation design and outcomes, and (2) to reflect upon the gaps and challenges of eHealth evaluation raised in previous research by using internal documents, collected project data (for example, patient data), and access to the stakeholders' reflections. To the best of my knowledge, no previous research has applied such an in-depth view. It should be noted that this research could not fulfill one aspect of a case study design; that is, a researcher "having little or no control" over the setting studied (Yin, 2014). Being actively involved in the project, I acknowledge this "disadvantage" in Chapter 3.5, where I reflect upon my role in the project and research.

3.2.2 Empirical setting

The empirical setting of this thesis (and all the included studies) was the three-year (2015–2018) European Union project "Digital Environment for Cognitive Inclusion" (DECI), which aimed to define and test an ICT-supported business model to provide digital services to elderly people with mild cognitive impairment (MCI) or mild dementia (MD). DECI introduced eHealth solutions and a supporting organizational model in four different healthcare contexts in Sweden, Italy, Spain, and Israel. The DECI solution included: (1) an integrated care platform to enable communication between patients, informal caregivers and care providers, and data sharing and storage; (2) a user activity monitoring system, which was an indoor sensor to monitor patients' activity; (3) a user coaching and training system, which provided a customizable physical training program to patients; and (4) a cognitive exercise system, which provided a customizable online cognitive stimulation program to patients. The role of a case manager was introduced in the care models to ensure the coordination and integration of different cervices into a cohesive program customized individually to meet the needs of patients

(Mueser et al., 1998). By integrating three domains of eHealth – direct provision of care, health information exchange and communication, and indoor sensors and wearables – DECI represented a large portion of the types of eHealth interventions.

Additionally, multinational and multidisciplinary nature of eHealth intervention exposed peculiarities and consequences of collaboration in the eHealth implementation projects that are implemented by a research consortium. Eight partners represented different business types, such as healthcare organizations, information technology firms, and science institutions (see Figure 3).

Figure 3. Organizations involved in the "Digital Environment for Cognitive Inclusion" (DECI) project

3.3 *Research process*

I joined the DECI project on day one of my PhD, in 2015. The project had concrete goals, deliverables, and a timeline of four years (2015–2018). As a Chalmers team, we were in charge of the summative evaluation of the DECI intervention. Within this context, my PhD took an inductive approach; at first, I immersed myself in the empirical setting, where I observed and actively participated in the practice of conducting an eHealth intervention in a multi-cultural and interdisciplinary setting. I participated in many project activities and gathered new experiences.

Being fresh to the field, I observed various phenomena during the project. After every key phase of the project, I discussed my observations with my supervisors through the potential research perspective. Curiosity then led me to check whether and how those phenomena were discussed in scientific literature. After problematizing the observations, I formulated research questions, formed collaborations for co-authorships, and pursued questions utilizing data connected to

DECI. This process applied to all the conducted studies throughout my PhD. It resulted in a longitudinal view on evaluation of DECI, from planning it till reflecting on the use of evaluation results. Figure 4 presents a timeline of data collection, analysis, and drafting of the manuscripts.

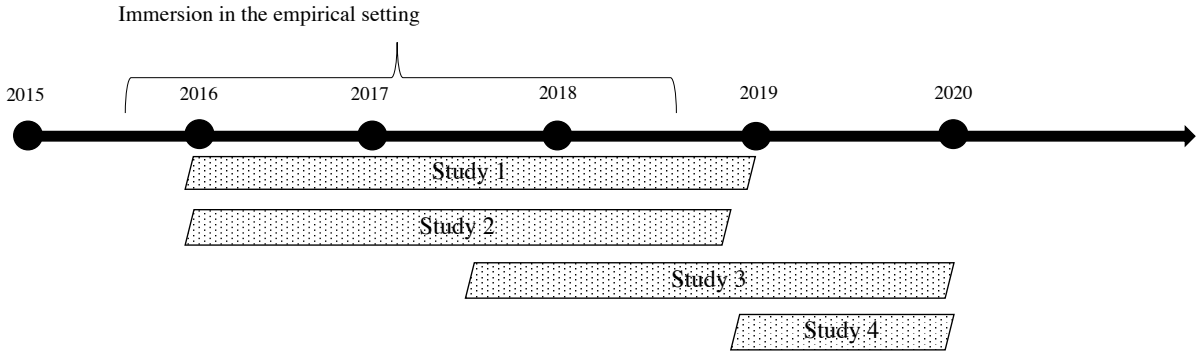

Figure 4. Research timeline

3.4 *Data collection and analysis*

This chapter describes the methods used in collecting and analyzing data for the studies included in this thesis. Table 1 presents the methods of data collection and analysis applied in the studies and papers.

Table 1. Methods of data collection and analysis per study and paper. '+' indicates that the method was applied.

Research question		**RQ1**		**RQ2**		**RQ3**
Study		**I**	**II**	**III**		**IV**
Paper		**1**	**2**	**3**	**4**	**5**
Data collection	Documents	+	+			
	Interview			+	+	+
	Survey			+	+	
Data analysis	Thematic analysis			+	+	+
	Pattern matching	+	+			
	Process coding	+				
	Quantitative analysis: - Wilcoxon test - Mann-Whitney U test - Descriptive statistics			+	+	

3.4.1 Mixed methods

A mixed methods approach (Greenhalgh and Russel, 2010; Dick et al., 2020) was implemented when studying the conceptualization and measuring of value of an eHealth solution (research

question 2). Therefore, a combination of qualitative and quantitative methods was applied. Qualitative research is particularly useful in understanding different meanings and motivations of individuals (Prasad, 1993). Hence, qualitative research methods (interviewing and document analysis) were used in studying Research Questions 1 and 3, specifically (1) understanding the outcomes of interactions between the project partners when planning eHealth evaluation, (2) exploring the views of project partners during planning and post-evaluation, and (3) finding out the perceptions of patients and healthcare professionals on the eHealth intervention.

Quantitative data came from a clinical trial that was part of DECI project. As is customary in medical research, quantitative data was collected through standardized surveys and it intends to provide a measure on the before-after condition of a patient. Additionally, cost data were collected using tailor-made surveys to the healthcare institutions involved in the project. However, while a survey is a 'snapshot' of a particular moment in time, the respondent is often guided by a pre-defined set of answer options, and this method does not allow the respondent to add his or her own interpretations. Such quantitative data was used in studying for Research Question 2 and provided a complementary view (Yin, 2006) on the value of eHealth intervention to patients.

3.4.2 Study I

Study I related to the first research question of this thesis and analyzed the adequacy of standard eHealth evaluation planning guidelines to facilitate the practical process and to ensure the planned evaluation is of good quality. Since no planning guideline was applied during the planning of DECI evaluation, the aim was twofold: (1) to study how the guidelines cover the essential events in practice of planning an evaluation, and (2) to reflect upon how practice could be improved if guidelines were used. Since these guidelines typically provide step-by-step guidance to evaluators, the study aimed to compare the empirical process of evaluation planning in DECI to the process defined in two most prominent guidelines for this purpose, namely Good Evaluation Practice in Health Informatics (GEP-HI) (Nykänen et al., 2011) and Health Information Technology Evaluation Toolkit (AHRQ) (Cusack et al., 2009).

Three out of four co-authors (including myself) were actively involved in planning the evaluation of DECI, but at the time of these activities, this study was not planned. This means that we analyzed our own (and other project partners') practice, which had already been completed at the time of the study, and compared it with the activities recommended in the standards. The study followed a qualitative research strategy and applied a process research perspective.

Data collection

For comparison between the planning process defined in the guidelines and the empirical process of planning an evaluation, Study I applied a process research approach (Langley, 1999; Langley and Tsoukas, 2017). It built on the historical data of how events unfolded throughout

the evaluation planning in DECI. Data set for the study included multiple sources of documents consisting of 262 e-mails exchanged between DECI project partners, eight meeting minutes, and 32 versions of evolving evaluation plan. Only the documents that fell into the evaluation planning period between September 2015 and September 2017 were included, until the evaluation plan was approved by DECI team and the project reviewers. These types of documents were deemed appropriate data for the study due to the fact that DECI partners were located in different countries and communicated mainly via e-mail and Skype calls. Therefore, this intensive collaboration was judged to be best reflected in these documents. Informal conversations that took place outside the formal meetings or e-mail communication were not included in the data set. Moreover, the method of considering e-mails as a data source has been shown valuable in similar studies that analyze the trajectory towards a common goal in large teams (Gehman et al., 2012). Since the Chalmers team was responsible for the evaluation activities in DECI, it can be assumed that majority of the documents were available for inclusion in this study (meaning that I was actively or passively involved in these communications). It should be mentioned that this responsibility by Chalmers also meant leading the planning activities (for example, organizing the calls and setting the agenda for discussions), which has affected how the evaluation planning process looked over time. However, other partners of DECI have also been actively involved and have influenced whether those activities took place and how they looked.

Data analysis

Because the primary focus was on events, we aimed to deduce the sequential activities (a process) (Langley, 1999; Langley and Tsoukas, 2017) in developing an evaluation plan of DECI. Therefore, a database of records, arranged chronologically, was initially created by combining data from all the sources (Gehman et al., 2012), resulting in 301 records. To avoid bias due to personal experiences being involved in the evaluation planning, a record was only included if it reflected in the data sources (Glick et al., 1990). A process coding technique was then used to organize the data (Saldaña, 2015). As the data were coded, patterns emerged (a pattern was considered a purpose of an activity reflecting in the data record), which later helped to identify the process steps. Afterwards, the codes were aggregated into categories based on the summative features (aggregated purposes) in the codes (Saldaña, 2015) and were eventually depicted as process steps. Sequence of the process steps was determined by the time stamps (Gehman et al., 2012). As the first author, I coded and categorized the data. As the interpretation of the data was subjective, I have thoroughly documented my reasoning and choices in the analytic memo (excerpts of it are presented in Annex 1), as recommended by Saldaña (2015). Two of the other authors examined the results of analysis and my reasoning, shared their insights, and we resolved discrepancies reaching the empirical process view that we felt most accurately reflected the reality of DECI evaluation planning.

In the second stage, we compared the DECI evaluation planning process with GEP-HI and AHRQ guidelines. A pattern-matching technique was used to compare between a theoretical

and observed patterns (Trochim, 1989). The patterns emerged through comparing the purpose and activities of a step. These were the factors for determining a match or no match between the empirical process and the guidelines. Steps in the guidelines and the empirical process that shared a similar purpose and activities were grouped as a "match," and those that had no similarities in activities or purpose were grouped as "no match." This analysis was performed using a tailor-made pattern-matching tool, which was essentially a matrix listing the process steps of DECI on one axis and the steps recommended by the guidelines on the second axis. This analysis was performed by three authors individually, and the results were compared and differences were resolved.

3.4.3 Study II

This study relates to the first research question of this thesis and explored what factors can hinder the use of standards in eHealth evaluations. Previous research has emphasized the usage of standards as one of the means to increase quality in evaluations (Proudfoot et al., 2011; Mookherji et al., 2015; Cowie et al., 2016). The paper was single-authored.

Data collection

The study built on the same documents as in Study I, but the focus was on discussions regarding the considered standards for evaluation between the DECI partners, as documented in e-mails (available to myself as a coordinator of evaluation) and meeting minutes dated between September 2015 and September 2017. Informal conversations that took place outside the formal meetings or e-mail communication were not included in the data set. Even though Chalmers had a leading role in the DECI evaluation planning, discussions regarding which standards should be included in the study took place mainly among clinical partners, and Chalmers' impact on those discussions was minimal (only to stimulate the discussion in order to reach a consensus).

Data analysis

Study II applied a process research approach building on the historical data of how events unfolded through the evaluation planning in DECI (Langley, 1999; Langley and Tsoukas, 2017). Events that took place when considering using a standard and hinders leading to different decisions in these events were of interest. At first, several of the standards used in DECI were selected for analysis in this study based on the opportunities to capture translation strategies (Røvik, 2007). The study included an eHealth evaluation framework, namely the Model for Assessment of Telemedicine Applications (MAST) (Kidholm et al., 2012), and three standardized measures that cover several evaluation themes of MAST (quality of life, patient satisfaction, and patient perspectives), namely the EuroQoL five-dimension questionnaire to assess health-related quality of life (EQ-5D-5L) (EuroQoL Group, 1990), the Patient Satisfaction Questionnaire (PSQ-18) (Marshall and Hays, 1994), and the Camberwell Assessment of Need for the Elderly – Short Form (CANE-S) (Reynolds et al., 2000; Orrell and

Geraldine, 2004).

Because the primary focus was on events, we aimed to deduce the sequential activities concerning each of the selected standards (Langley, 1999; Langley and Tsoukas, 2017). Data reflecting events and discussions on usage of each standard were outlined in chronological order using time stamps in the data (Gehman et al., 2012; Saldaña, 2015). The content of the events was examined to reveal how the decisions of using the selected evaluation standards evolved in a multinational and interdisciplinary team. This helped us identify the evolution of events for each standard, from an idea to use the standard to the decision on its actual use (Gehman et al., 2012).

The data were organized using the process framework of the development of cooperative inter-organizational relationships (Ring and Van de Ven, 1994). Events and quotations by project partners were assigned to the categories of the framework, namely negotiation, re-negotiation, commitment (agreements), and execution. Then, decisions on the use of standards were interpreted as translation strategies, as defined by Røvik (2007). A standard was anticipated to be translated using *copying* strategy if the standard was used with no modifications to its content; by *addition* strategy if the content of a standard was supplemented by extra elements; by *omission* strategy if one or more components were removed from the content of a standard, for certain reasons; and by *alteration* strategy if the standard was altered to a large extent, but not in the same way as omission and addition strategies. Finally, factors hindering the use of evaluation standards were identified in the project partners' quotations embedded in the events.

The data analysis was performed by me (a single author of the paper), but the "stories" regarding each standard were presented to three supervisors for questioning regarding the choice of the standard and the stories' ability to capture the translation strategies. In addition, the "stories" and the draft of the paper were sent to DECI partners (those whose quotations formed the data set) for verification regarding the accuracy of the events and for allowing any extra data to be uncovered.

3.4.4 Study III

This study relates to the second research question of this thesis. It aimed to elaborate on the interpretation of the 'value' concept by reflecting upon the created value of an eHealth service in different contexts, namely Sweden and Italy. To address this question, the study aimed to understand the uniqueness of value that an eHealth intervention delivered by comparing outcomes in two different contexts. To organize the outcomes, a conceptual model for the value of eHealth intervention was proposed in terms of monetary and non-monetary benefits and sacrifices, inspired by earlier research (Ramirez, 1999; Grönroos, 2008; Gummerus, 2013). The model was populated with variables that were commonly used in regular clinical or organizational practice by DECI partners and that were feasible for collecting data on given the time and resource constraints of the evaluation assignment. Based on these variables, the

monetary benefits were operationalized as income and prevented cost of treatment. The non-monetary benefits included clinical efficacy, quality of life, patient satisfaction, and job satisfaction. The monetary sacrifices were operationalized as investment, operating expenses, and cost of spent time. The non-monetary sacrifices were expressed as patient safety and workload. The data collection and analysis for those domains is explained below.

Data collection

Tailor-made surveys (a Microsoft Excel file) were distributed to the contact persons from Italian and Swedish clinical sites at the end of the clinical trial to supply the quantitative data on income (reimbursement rates) and cost (such as hourly rates, treatment tariffs). To enhance scientific rigor, standardized questionnaires were used to collect patient data in DECI (Mookherji et al., 2015). These included mini-mental state examination (Folstein et al., 1975), clock-drawing test (Shulman, 2000), EQ-5D-5L (EuroQol Group, 1990), and Clinical Dementia Rating Scale (Morris et al., 1997). The questionnaires were used with patients at the baseline and after six months of the intervention. These data using the questionnaires were collected by the healthcare professionals. Semi-structured interviews concerned the user experience on DECI, impact on the patients and healthcare professionals' work routines. This made it possible to obtain the necessary information and to deviate following the respondent's line of reasoning through follow-up questions (Kvale, 2007). Based on the pre-agreed interview protocol within the research consortia, patients were interviewed by the healthcare professionals during the last visit in the study. Healthcare professionals were the interviewers because the aim was to optimize resources in the project and to reduce cognitive load for patients, which would have potentially been incurred if they were contacted by the unknown people to them. The healthcare professionals involved in the intervention were interviewed by the Chalmers team. Where data was necessary for estimating the value dimensions, but not collected in DECI, those data were theorized based on the relevant literature that built on the same eHealth solution or similar population. For example, data on falls by elderly people were not collected in DECI, but such data were necessary for calculating the prevented cost of treatment. Therefore, the data of the average falls prevention rate for elderly were borrowed from the literature that built on a DECI's physical exercise program OTAGO, which had also been used earlier in another project and the results of which had been published.

Data analysis

Statistical analyses were conducted for determining changes from baseline to six months in patient quality of life and clinical efficacy using SPSS, Wilcoxon test. Between-group differences were calculated using Mann-Whitney U test. Monetary benefits and costs were calculated by multiplying (1) cost for a particular single treatment to the annual targeted patient population, (2) an hourly rate of a healthcare professional occupation to the number of hours spent per year for a particular treatment throughout six months of the study and aggregated to a yearly cost, and (3) various technology fees were projected to the yearly cost. A thematic analysis was conducted on the interview data to generate insights on the patient and healthcare

professional perspectives (regarding safety, workload, satisfaction), and to compare and highlight differences in these data (Braun and Clarke, 2006) based on countries (Sweden, Italy) or user types (patients, professionals). Patient interview data were quite brief, mostly due to the specifics of dementia and because the interviewers (healthcare professionals) aimed to avoid overloading the patients with information. An inspiration for the themes was also searched in the literature, particularly the technology acceptance model (Venkatesh et al., 2003), the innovation diffusion model (Rogers, 2003), and service quality (Parasuraman et al., 1988), as these models could, to a large extent, cover the aspects of eHealth that encompass technology, innovation, and service management (Kaplan and Shaw, 2004). The themes were selected from the models based on the assessment of whether such data were available in the collected data set or not. Moving within the data set and checking with the aforementioned models, the final themes included "workload", "relative advantage", "appreciated features of technology", "non-appreciated features of technology", and "aesthetics". The last step was to identify similarities and differences between different countries' respondents within the themes. Finally, a monetary benefit-sacrifice ratio was calculated for Years 1, 2, and 3.

3.4.5 Study IV

This study relates to the third research question and explored how different stakeholders used evidence from an eHealth trial. The aim included identifying the barriers that prevent stakeholders from using the evidence.

Data collection

The study took a qualitative approach. I conducted nine semi-structured interviews with the partners from DECI research consortium, representing four care centers, two research and development companies that provided the eHealth solutions for the trial, and two science institutions. The goals of the interviews were to understand the attitudes of the stakeholders regarding the evidence generated by DECI and to learn the plans and the actions that had already been taken in relation to the evidence. The interview questions included such topics as stakeholders' agenda and achievement in the project, how the evidence from the project was used, what evidence were used, outcomes and learning achieved as a result of using the evidence, communication of the evidence, timeliness of the evidence for the organization, and future plans with the evidence (Alkin and Taut, 2002; Leviton and Hughes, 2016; Alkin and King, 2017). Using semi-structured interviews made it possible to obtain this information and to deviate following the respondent's line of reasoning through follow-up questions (Kvale, 2007). The interviews were conducted via Skype and lasted for 1 hour. I transcribed the interviews verbatim.

Data analysis

I analyzed the interview data using thematic analysis (Braun and Clarke, 2006), with the aim of analyzing particular aspects of evidence use (the same aspects were incorporated in the interview questions) (Alkin and Taut, 2002; Leviton and Hughes, 2016; Alkin and King, 2017).

First, the data were organized based on the types of evidence used by a stakeholder (evaluation results, experiences, or previous research) (Evidence-Based Medicine Working Group, 1992); that is, which stakeholders had made use of which type of evidence. Then, the stakeholders' goals for using every evidence type were identified (their reasoning and agendas before the project and after the evidence were obtained). Finally, the ways of using the evidence were determined (instrumental, conceptual, or symbolic) (Alkin and Taut, 2002; Leviton and Hughes, 2016; Alkin and King, 2017) based on the stakeholders' actions taken or results achieved using the evidence. Finally, the entire data set was reviewed once more to identify the barriers that prevented the stakeholders from using the evidence. The analysis has been reviewed by the co-author in this study. Discrepancies in interpretations were discussed and resolved.

3.5 *Reflections on my background and role in DECI and its influence on research*

Prior to becoming a PhD student, I was a process management professional in various industries, primarily focusing on business process improvement. I also used to work in the management consulting field, where I helped clients to develop quality management systems based on the ISO9001 standard. I also hold a lead auditor certificate for ISO9001. I had no previous experience in eHealth or evaluation, although I was knowledgeable of the specifics of EU projects from my former employments. I believe that my background shaped my research in a number of ways. First, I took a longitudinal view of evaluation, from planning till utilizing its results trying to grasp the processes surrounding the evaluation. Second, several methodological choices were based on the process research perspective and manifested through searching for sequences in activities and determining of how events unfold or comply with methodological documents (resembling auditing). Third, my previous work with process improvements frequently focused on reducing waste in a broad sense (resource, time, effort). Due to this skill of observing areas of waste, the problem of pilotism (Tomlinson et al., 2013; Andreassen et al., 2015; Urueña et al., 2016) resonated with me and I became curious to understand what contributes to this problem in the context of eHealth evaluations. For the same reason, I took action to achieve consensus regarding what to measure among the DECI stakeholders, aiming to increase chances that evaluation evidence would be used after the study ends.

Chalmers was in charge of the DECI evaluation. Activities and results of this work served as data for my research, but to some extent these data were shaped partly by my own actions (from a methodological perspective) in the project. Being an inside researcher meant taking the role of coordinator of evaluation activities as well simultaneously becoming a member of a research community outside the project. However, insiderness was essential in order to experience the internal team-related processes of evaluation (for example, evaluation planning or selecting the evaluation methodology) and to observe interaction between different stakeholders. My practical experience in the project revealed the tension between theory and practice in eHealth

evaluation and raised curiosity that led me to formulate and pursue research questions that would not be accessible for outsiders. There is a chance that insiderness could create bias in data analysis (such as coding), and I have noticed myself checking whether I possess data to support the observed story I want to tell (as is probably natural in the research that builds on historical data). To increase transparency of the data analysis, I have carefully documented the analytical procedures and decisions. These have been revised by my co-authors, discussed, and disagreements have been resolved.

These aspects influenced my attempts to enhance research quality described in Chapter 3.6 below.

3.6 *Research quality and ethics*

Guba and Lincoln (1989) provided a set of quality criteria for research consisting of credibility, transferability, dependability, and confirmability. In addition, Leung (2015) discussed criteria of validity and reliability.

Credibility

Credibility is a criterion understood as a "value of the truth" (Halldórsson and Aastrup, 2003). Erlandson et al. (1993) stated that credibility of research depends on how a researcher has attempted to increase objectivity of the presented evidence due to limitations caused by his/her own construction of reality. Erlandson also emphasized that no single reality exists, and a researcher must be aware of his/her own constructions (Halldórsson and Aastrup, 2003). Table 2 introduces actions taken to increase credibility in the conducted research.

Table 2. Actions to increase credibility of the research

Study I	The "own constructions" in the development of the DECI evaluation planning process and its comparison with the eHealth evaluation planning guidelines (Paper 1) were validated by triangulating the understanding by 'multiple observers' (Denzin, 1970). In the present study, these were three researchers performing the same analysis and then comparing the results and discussing and resolving discrepancies.
Study II	The use of quotes from e-mails by project partners was validated with the partners by sharing with them the written paper and receiving a consent that the quotes were used properly and that they reflected the "truth" (Diener and Crandall, 1978; Bryman and Bell, 2011).
Study III	Thematic analysis of semi-structured interviews was based on the theoretical propositions from highly diffused models that explain various aspects of using eHealth, in order to avoid "own constructions" and to utilize the models (Yin, 2015).
Study IV	The use of quotes from interviews with project partners was validated with the partners by sharing with them the written paper and receiving a consent that the

	quotes were used properly and that they reflected the "truth" (Diener and Crandall, 1978; Bryman and Bell, 2011).

Transferability

The transferability aspect of the research quality refers to the relevance of the findings to other settings (Halldórsson and Aastrup, 2003). The findings generated depended on the empirical context that was related to a summative eHealth evaluation involving a multinational and multidisciplinary collaboration. With the growth of such collaborations in the eHealth research (Greenhalgh and Russel, 2010), the results of this thesis can provide reference for analytical generalization (Yin, 1989). In addition, transferability was increased by using standardized instruments for outcome measures, so that the research results were more trustable and could be used for comparisons in other settings (Lilford et al., 2009; Mookherji et al., 2015; Cowie et al., 2016).

Dependability

Dependability relates to the "audit trail" of the conducted research (Guba and Lincoln, 1994). Dependability emphasizes the importance of being able to track the research process, method, and decisions made. Table 3 introduces to actions taken to increase dependability in the conducted research.

Table 3. Actions to increase dependability of the research

Study I	Data and analytical reasoning during the research process were documented in detail in an analytic memo, and decisions were registered in the minutes of the meeting (Saldaña, 2015). The files documenting the analysis and reflections on the DECI process of evaluation planning and the theoretical guidelines were stored.
Study II	Relevant quotes from the e-mails reflecting the discussions related to the use of different standards were extracted from the e-mails that were stored in the IT server and available for retrieval, if needed.
Study III	Data and analytic files containing calculations regarding the benefits and sacrifices of using an eHealth solution were stored in the IT server and are available for retrieval, if needed.
Study IV	Relevant quotes from the interviews reflecting the use of evidence from the eHealth intervention were extracted from the interview transcripts that were stored in the IT server and available for retrieval, if needed.

Confirmability

Confirmability of the research concerns a potential bias of a researcher and limitations affecting the research process and findings (Halldórsson and Aastrup, 2003). Therefore, reflecting upon one's own values and standpoint in relation to the research is advisable (Bryman and Bell, 2011). The methodological choice made in Studies I and II to use data from e-mail communications has potential limitations. For example, I might not possess all e-mails that had circulated among project partners in relation to the DECI evaluation planning. However, this would be less likely to occur since I was in charge of the evaluation activities. In Study II, the choice of standards to include in the analysis was based on personal observations and richness of data in my possession. A different set of standards might have provided different results. However, this can be a future research possibility.

Some of the interviewees had a dual role as healthcare professionals who were interviewed about the DECI impact on the patients and work routines (Study III), and also as the key representatives of the project partners interviewed about the post-project activities related to the collected evidence (Study IV). However, this duality was resolved by using two different interview protocols with non-overlapping questions and an introduction during every interview explaining the purpose of the interviews.

Validity

Validity addresses the suitability of the chosen sample, measures, and methodology in answering the research question, and how the results are valid for the sample and context (Leung, 2015). Table 4 introduces actions taken to increase validity in the conducted research.

Table 4. Actions to increase validity of the research

Study I	E-mails and other documents reflecting series of events that took place during the evaluation planning were considered as a valid data source for understanding how events concerning eHealth evaluation planning unfold over time since they have been used in similar studies that analyzed the trajectory towards a common goal in large teams (Gehman et al., 2012). A single-case method design also allowed for an in-depth exploration of the events (Yin, 2015), which was necessary for a longitudinal view on events. The results reflected the events that might take place in the multidisciplinary and international research set-ups.
Study II	E-mails and other documents containing decisions made over time within the research consortia were considered as a valid data source for understanding the problems surrounding the selection of the standards for eHealth evaluation since it has been used in similar studies that analyze the trajectory towards a common goal (agreeing on a common standard for use) in large teams (Gehman et al., 2012). A single-case method design also allowed for an in-depth exploration of the events (Yin, 2015), which was necessary for a longitudinal view on events. The results reflected the events that might take place in the multidisciplinary and international research set-ups.
Study III	The study aimed to identify the outcomes of an eHealth intervention. Only validated outcome measures, standard in clinical practice, were chosen for the study. Qualitative analysis of semi-structured interviews was based on the

	theoretical propositions from highly diffused models that explain various aspects of using eHealth, avoiding "own constructions" and utilizing the models (Yin, 2015). Validity of the research design was aimed from the medical research perspective, and patient outcomes were compared in the before–after situations of a patient and differences between the comparison groups. However, from the social science perspective, validity of such a research design is limited due to other confounding factors that play a role in a non-controlled home environment of a patient; a concern discussed by social researchers (Andargoli et al., 2017; Dick et al., 2020).
Study IV	A semi-structured interview was a suitable method for the study, since the research question required to obtain an overview of the plans and actions already taken by the limited sample of partners involved in the concrete project. Semi-structured interview allowed for discussions beyond the research protocol, helping to understand the reasons behind the decisions made.

Reliability

Reliability addresses the consistency in process and results of research. It can be achieved through comprehensive data use, constant data comparison, checking accuracy of the data extracted from other sources, and triangulating data (Leung, 2015). Table 5 introduces the actions taken to increase reliability in the conducted research.

Table 5. Actions to increase reliability of the research

Study I	For consistency in the analytic process, analytic procedures and rationale for the decisions made were thoroughly documented during the analysis (Saldaña, 2015). Three researchers employed the pattern matching technique (Trochim, 1989) in the independent analysis and then compared the perspectives for achieving a common view of how events unfolded.
Study II	Data were obtained from two different sources (e-mails and meeting minutes). For consistency in the analytic process, time stamps in the data were used for outlining a chronology of events. In these data, triangulation of the perspectives by different DECI partners for using a standard allowed to obtain a multi-angled perspective on why achieving a consensus regarding the use of a standard might be complicated in the inter-disciplinary and multi-national research set-ups.
Study III	Data accuracy was achieved through logging the research data into a digital platform. The data that were extracted for analysis contained time logs that showed at which point in the intervention the data were recorded (six months were needed for the before–after situation comparisons). During the data cleaning procedures, the time logs were checked for inclusion of the data in the analysis. Afterwards, accuracy of the data was also checked by carefully reviewing the data sample and eliminating data that had the wrong format.
Study IV	Where possible, several interviewees from the same organization were involved in the same or separate interviews. Triangulating their perspectives made it possible to obtain a more accurate picture of the evidence used in that

	organization.

Ethics

First, the DECI project study protocol has undergone ethical approval procedures in all the project locations (Sweden, Italy, Spain, and Israel) that were the pilot sites of the intervention. Second, Diener and Crandall (1978) suggested four main areas for potential ethical issues in business research: harm to participants, lack of informed consent, invasion of privacy, and deception.

(1) *Harm to patients* participating in DECI was considered in terms of physical harm and psychological stress due to the intervention or data collection processes. Physical harm was avoided in two ways: first, by baseline measurements assessing patients' health state before starting the intervention (Lilford et al., 2009), and second, the physical exercises were personalized based on the age and health state of a person. Stress during data collection was addressed through a limited number of interview questions and convenience sampling (Palinkas et al., 2015) where only patients capable of participating in the interviews were included. Indirect harm could also be caused by "taking away" the eHealth solution or a new way of care from the patients once the intervention was over, especially when the qualitative interviews had indicated appreciation of the new ways of care by the patients. After the intervention, most clinical sites of DECI took action and consideration of whether and how the DECI model could be made available to the patients on daily basis (Paper 5).

(2) *Harm to stakeholders* participating in Studies II and IV concerned the potential harm to their careers due to the e-mail content and interview questions that exposed internal dynamics within the project consortia and required reflections upon the evaluation process and the use of evidence. The harm was avoided by providing the stakeholders whose quotations have been used in the papers with the possibility to review the final versions of the papers before submitting. E-mail consent to publish was obtained from each individual stakeholder.

(3) *Informed consent* can provide legal protection to the investigator in case of a harm to a patient during a study (Erikson, 1967). However, it becomes critical to provide full information regarding the study and its risks (Homan, 1991). In DECI, the informed consent was signed by every patient who participated in DECI. In addition, information about the study was provided in an individual meeting with each patient and its caregiver.

(4) *Invasion of privacy* relates to a disclosure of personal information that leads to the possibility of tracking a person's identity (Bryman and Bell, 2011). After entering the patients' data into the DECI system for data collection, every patient received a unique ID that only disclosed age and gender. When it comes to the interviewed stakeholders,

their identities were hidden, to a certain point (Paper 2). Although names were avoided, the journal demanded that information be disclosed regarding interviewees' occupation, gender, and country. The stakeholders agreed to have this information disclosed in the paper before the paper was published.

(5) *Deception* refers to hiding the true purposes of the research from its participants (Bryman and Bell, 2011). In DECI, detailed information about the study was provided in an individual meeting with each patient and its caregiver. When it comes to the interviewed stakeholders, the purpose of research was explained when scheduling the interviews and the paper was shared with every interviewee before publication.

4 SUMMARY OF APPENDED PAPERS

This chapter provides a brief introduction to the results and contributions of each five papers appended in the thesis, in relation to the purpose and research questions of the thesis.

4.1 *Paper 1: 'Planning a holistic summative eHealth evaluation in an interdisciplinary and multi-national setting: A case study and propositions for guideline development'*

Previous research has emphasized that quality in eHealth evaluation studies is lacking, which reduces the value of the results. Evaluation quality problems include application of non-scientific methods and non-standardized instruments for measurement, insufficient attention to- and unfeasible scope of data collection, mismatches between outcomes measures and technology, or false assumptions about data quality. The purpose of Paper 1 was to explore adequacy of eHealth evaluation planning guidelines to support quality in evaluations.

The studied case adds aspects that could help prevent quality problems in eHealth evaluations that are not covered in the current evaluation planning guidelines. First, the guidelines could add a recommendation to embed an evaluation in a context of previous research and comparable studies, and to check what evaluation methods had been used previously. Not covering this aspect can hinder continuity and learning. Second, the guidelines overlook the importance of monitoring data quality, taking the laws of data protection and privacy into account, considering ethical and legal aspects during the evaluation planning, and performing a risk analysis for the evaluation process. Also, the guidelines could address and promote collaboration in various set-ups of stakeholders by considering processes of social dynamics, such as negotiations, promoting own agendas, and consensus seeking, especially observed when trying to align evaluation methodology. The paper suggests including these domains in the current guidelines for eHealth evaluation planning.

Paper 1 contributes to previous research by enhancing current evaluation frameworks with new domains that can help prevent quality issues in eHealth evaluation studies and consequently increase the value of the evaluation results and evidence created. In addition, the multinational and interdisciplinary nature of the empirical setting offers a broader perspective than that of prior frameworks.

4.2 *Paper 2: 'Standards as applied in reality: a case study on the translation of standards in eHealth evaluation practice'*

Previous research has emphasized that scientific rigor is key for producing quality evidence in a summative eHealth evaluation. Applying evaluation standards can strengthen trust and

promote usage of evidence. However, it has been recognized that the use of scientifically valid evaluation standards was lacking in the eHealth evaluation studies. One reason could be that usage of standards is voluntary. Another reason relates to the evaluation set-ups. For example, interdisciplinary evaluations can be burdened by contesting agendas and practices. The purposes of this paper were to explore how standards are used in a practical setting of an eHealth evaluation and to identify the factors that can hinder their use.

Findings from the studied case indicate that the collaborative processes and events during evaluation planning differ depending on the type of the standard. It is easier to agree about using a generic evaluation framework than about specific standardized measures that caused extensive negotiations. In addition, a number of hindering factors were identified, namely, (1) insufficient fit of a standard to address a target population or a disease, (2) lack of resources to use a standard for data collection, (3) lack of experience in using a standard, and (4) non-existent validated versions of a standard in a particular location.

This paper contributes to earlier research by proposing an understanding of how practitioners choose, modify, or reject the use of evaluation standards creating varying degrees of adherence to the original content of the standard. This creates heterogeneity of methodologies among different eHealth evaluation studies. Consequently, modifications in standards can affect the trust in evidence and its comparability across different studies. Based on the identified barriers to the use of standards, necessary means and guidance can be further developed to reduce the risk that standards would not be used, thus affecting the quality of evaluations.

4.3 *Paper 3: 'An Italian business case for an eHealth platform to provide remote monitoring and coaching services for elderly with mild cognitive impairment and mild dementia'*

When making investment decisions to adopt an eHealth solution in a healthcare system, decision makers seek to maximize the return on investment by assessing the business case of using a technology. Existing research is scarce when it comes to the published cost and benefit evaluations for eHealth solutions targeted at cognitive impairments, especially regarding the economic data. This paper aimed to describe an Italian business case of using an eHealth platform to provide remote monitoring and coaching services for elderly people with mild cognitive impairment and mild dementia.

By assessing various monetary and non-monetary outcomes, the findings indicate positive health outcomes (postponing care need and saving cost due to delayed hospitalization) and increased patient satisfaction and healthcare staff satisfaction compared to the usual care. Using the eHealth solution requires more financial resources than the usual care. However, the new care model saves on the cost of hospitalization and the treatment can be partially reimbursed by the state. This creates a situation whereby the monetary side becomes positive, meaning that

the care can be improved at no financial loss by the healthcare provider.

The paper contributes to earlier research by providing evidence on a positive business case of an eHealth solution to support patients with cognitive impairments. Decision makers can use this evidence for benchmarking other eHealth solutions when making investment decisions.

4.4 *Paper 4: 'Identifying the Value of an eHealth Intervention Aimed at Cognitive Impairments: Observational Study in Different Contexts and Service Models'*

While the concept of 'value' has been used in healthcare, it is not well defined within the context of a summative eHealth evaluation. When discussing value, scholars have referred to clinical efficacy, behavioral change, perception-based feedback, or economic terms of implementing or testing an eHealth solution. These different focus areas show that there has been no conceptualization on how to evaluate value in eHealth interventions. In other fields, value has been perceived as benefits and sacrifices, and relativity of this concept has been emphasized. This multiplicity of interpretations not only creates confusion for interpreting eHealth evaluations, but also creates premises to question the potential of learning and continuity that evidence produced in evaluation conducted in a particular context can provide. The purposes of this paper were (1) to identify the contextual factors that determined the similarities and differences in the value of an eHealth intervention between two different contexts, and (2) to reflect on and contribute to the discussion about the specification, assessment, and relativity of the "value" concept in the evaluation of eHealth interventions.

The findings indicate that the same eHealth intervention introduced in different contexts delivers different value to its users, both economically and in terms of non-economic aspects. In one context, clinical outcomes and patient satisfaction were improved, at the expense of an increased staff workload and substantially higher cost. In the second context, stability of the clinical condition improved, satisfaction of care by staff and patients increased, and the workload was acceptable. Besides these outcomes, monetary benefits outweighed monetary costs in this context from the first year of using the eHealth solution. Factors that influenced these differences in value were (1) the service delivery design (process) of the intervention, (2) the organizational arrangement of the intervention, (3) the cost of different treatments, (4) the hourly rates of staff for delivering the intervention, and (5) the lifestyle habits of the population.

The paper contributes to earlier research by proposing a conceptual model for determining value of eHealth interventions in a particular context. The model includes monetary and non-monetary benefits and sacrifices. By utilizing this model to evaluate an eHealth intervention in different contexts, the paper demonstrates that differences in value manifest when a holistic view is applied. Results from such an evaluation can serve as a basis for modifying real-life applications in order to increase the value in a particular context.

4.5 *Paper 5: 'Exploring the Use of Evidence From the Development and Evaluation of an Electronic Health (eHealth) Trial: Case Study'*

Evidence-based approaches are increasingly being applied in evaluating eHealth solutions. Evidence (patient preferences, professional opinions, evaluation outcomes, and existing scientific evidence) should be generated through rigorous studies and used in decision making regarding eHealth integration into clinical practice and policymaking. However, there are concerns that the use of evidence is lacking in the context of eHealth, which causes a problem of 'pilotism', when pilot projects do not result in tangible and sustained improvements. Most previous research that has followed these arguments has focused on the evaluation outcomes, and the use of other types of evidence (such as patient preferences and professional opinions) is underexplored. In addition, extant research has mostly referred to the instrumental use of evidence in making decisions for practice change and policymaking. Previous studies have also neglected the multi-disciplinary nature of the eHealth field when discussing the beneficiaries of evidence, delimiting them to healthcare professionals and policy makers. Therefore, the purpose of the present paper was to analyze how various stakeholders use different types of evidence (such as evaluation outcomes including patient preferences, professional experiences obtained within the development and evaluation of an eHealth trial, and existing scientific evidence from other research). An additional aim was to identify barriers to the use of evidence and ways to support its use.

The findings indicate that evidence from eHealth trials can be used in a range of activities, such as scientific publishing and dissemination, eHealth technology improvement, research funding applications, and teaching. Professional experiences are more impactful than evaluation evidence when it comes to making eHealth adoption decisions in healthcare systems, and contribute greatly to policymaking, teaching, and improving technology. Therefore, it is important to create conditions for healthcare professionals to be involved in eHealth interventions. Evaluation evidence is mostly relevant to scientific publishing. Previous research can also complement eHealth adoption decisions if the context is relevant. There were two barriers to the use of evidence. The first was poor quality of some types of evidence caused by healthcare professionals' failure to collect all the necessary evidence due to the overly large scope of variables to collect. The second was evaluation design comparing before-after situations that did not allow for adjustments and learning along the intervention, making evidence about the eHealth solution's fit to the local context less relevant.

Paper 5 contributes to previous research by exemplifying a broader range of uses of evidence from eHealth trials when looking at different stakeholders and suggests that the concerns related to the 'waste' of evidence might be inflated. These findings should help to set more realistic expectations regarding what benefits can be brought by evidence in the form of an evaluation outcomes, professional experiences, or previous research. The paper also provides evidence that hard facts (or evaluation evidence) play a smaller role in decisions in the organizations than what is assumed in the scholarly community. The paper lists a number of strategies that could

support the use of evidence in eHealth context. First, it is advisable to refrain from extensive scope of variables in the evaluation. Instead, a few variables create conditions for better quality of data collected. Second, evaluation designs that enable iteration, learning, and creating experiences could provide more value to different stakeholders.

4.6 *Contributions to improving quality in eHealth evaluations*

Based on the core findings in this thesis and eHealth evaluation criteria outlined in the previous research, an extended view of eHealth evaluation quality is presented (see Table 6, which introduces this view on eHealth evaluation quality).

Table 6. The extended view on eHealth evaluation quality

Criteria of quality in eHealth evaluations	***Description***	***Extended view on eHealth evaluation quality, added by this thesis***
Contextual sensitivity	'Context' refers to a combination of economic, political, demographic, technological, organizational, stakeholder network, involved individuals' related factors (Greenhalgh and Russel, 2010; Andargoli et al., 2017; Dick et al., 2020). Contextual features provide background to the studied phenomenon, help to make sense of the findings, and help shape practical actions after the study (Greenhalgh and Russel, 2010).	*The same eHealth solution implemented in different contexts can bring different value (Paper 4). Therefore, summative eHealth evaluations need to account for contextual factors.* *Stakeholders need 'local' evidence sensitive to a particular context for activities like adopting an eHealth solution, policy making, and teaching students. In these activities, evidence produced in other contexts is less useful (Paper 5).*
Multi-stakeholder perspectives are considered	Summative eHealth evaluations can be affected by contesting stakeholders' needs, concerns, relationships, values, beliefs, and agendas. It takes time to develop a common 'language' (Catwell and Sheikh, 2009; Greenhalgh and Russel, 2010).	When planning a summative evaluation, it is worthwhile gathering the needs for, or plans with, evidence based on agendas of the stakeholders. Quality in evaluations increases when more stakeholders find evidence useful for their purposes, from learning to making investment decisions. Such information can be valuable input to the evaluation design (Paper 1, Paper 5). The feasibility of collecting the data for multiple stakeholders should be considered in light of resources at hand. Overambitious scope of outcomes measures can jeopardize the quality of the collected data (Paper 5). At the end of evaluation, it is recommended to conduct reflective activities, reserving time to discuss lessons learned (including methodological ones) and value of evidence produced (Paper 5).
Adequate	RCT remains the preferred	The choice of evaluation design needs to be informed

Criteria of quality in eHealth evaluations	***Description***	***Extended view on eHealth evaluation quality, added by this thesis***
research design	research design by many evaluators, and it is well suited in the high-risk eHealth interventions (Dick et al., 2020). However, arguments against RCT for lower risk eHealth interventions (Kaplan, 2001; Gurman et al., 2012; Nilsen et al., 2012; Mookherji et al., 2015; White et al., 2016; Pham et al., 2016) have led to a number of alternative research designs to choose from (e.g., Collins et al., 2005; Collins et al., 2007; Catwell and Sheikh, 2009; Greenhalgh and Russel, 2010; Mohr et al., 2013; Klasnja et al., 2015).	by the needs for evidence and knowledge regarding which evidence 'counts' for the stakeholders. Evaluation design might prevent evidence from being used (Paper 2, Paper 5). Evaluation design should allow healthcare professionals to form experiences with an eHealth solution (Paper 5).
Methodological pluralism	Mixed methods evaluation is the recommended approach during a summative evaluation, making it possible to obtain a snapshot of outcomes and reveal the social complexities of an intervention explaining the outcomes (Greenhalgh and Russel, 2010; Dick et al., 2020).	The importance of employing qualitative research methods during a summative evaluation is highlighted by a vital role that the captured users' experiences play in decisions of adopting an eHealth solution (Paper 5).
Scientific rigor through adherence to standards	Usage of standards leads to various degrees of methodological uniformity between different studies (including cross-cultural) and enhances generalizability and trust in the research findings (Bates and Wright, 2009; Mookherji et al., 2015; Cowie et al., 2016; Dick et al., 2020).	Adherence to a standard must be understood as a range, not a binary scale (adherence or no adherence). By making trade-offs between standards' content and circumstances, evaluators compromise adherence for a better contextual fit of a standard (Paper 2). More cross-cultural validations of instruments for measuring outcomes need to be conducted to make standards available for use across contexts. This would support methodological uniformity among studies, generalizability, and learning (Paper 2).
Ethical handling of data	Data are captured, structured, analyzed, managed, and shared in an ethical way (Mechael et al., 2019).	-
New criteria for quality added by this thesis: *Capturing value of an eHealth solution*	Different eHealth evaluation frameworks outlining themes of evaluation exist (e.g., Kidholm et al., 2012; van Gemert-Pijnen et al., 2011; Andargoli et al., 2017) without establishing an overarching conceptual purpose for assessment and how to make	The thesis conceptualizes value as a holistic view of the created monetary and nonmonetary benefits of eHealth that require monetary and nonmonetary sacrifices in a particular context (Paper 4). The model for value assessment of an eHealth intervention is proposed (Paper 4). In the model, evaluation domains populated with evaluation themes

Criteria of quality in eHealth evaluations	*Description*	*Extended view on eHealth evaluation quality, added by this thesis*
	sense of disparate evaluation themes.	need to be contrasted against each other, from the perspectives of multiple stakeholders. The scope of evaluation themes should be feasible and not jeopardize the quality of data collection and eventually the produced evidence (Paper 5).
New criteria for quality added by this thesis: *Involving healthcare professionals in the intervention and its evaluation*	When evaluators actively engage in the intervention and evaluation activities, they gain deeper insights into the studied phenomenon (Greenhalgh and Russel, 2009).	In addition to evaluators, it is worthwhile engaging healthcare professionals in data collection and evaluation activities, giving them opportunities to be analytical and reflective on (1) an aggregated level (beyond patient-by-patient basis), and (2) their own experiences. Such an approach makes them producers of summative evidence, not consumers, and can increase their involvement in using evidence for learning or improving care practice. In such a case, accountability and non-bias can be achieved through collaborations with external evaluators, revision of data and analysis (Paper 5).

5 DISCUSSION

The purpose of this thesis was to study how summative eHealth evaluations can be improved to support the determination of eHealth value in a specific context as well as the use of evidence produced during eHealth interventions. The following chapters discuss research outcomes for the research questions of this thesis. Finally, an extended list of criteria for eHealth evaluation quality is outlined.

5.1 *RQ1 – How can application of eHealth evaluation standards be improved?*

Standards represent aggregated, best-practice knowledge in the field (Brunsson and Jacobsson, 2000). In the context of eHealth, it has been argued that failure to apply standards can compromise the quality and scientific rigor of eHealth evaluations (Proudfoot et al., 2011; Mookherji et al., 2015; Dick et al., 2020). However, there is a gap between standardization in eHealth evaluations and the practice of evaluating. Research described in this thesis shows that standards diffused in medical research do not fully translate in eHealth evaluations compromising its quality (Paper 1 and Paper 2). For example, the majority of the standard metrics, such as questionnaires and scales (e.g., Marshall and Hays, 1994; Reynolds et al., 2000) that are widely used in eHealth evaluations originate from medical research. Moreover, there are multiple available standards for the same medical area, and evaluators need experience when using a particular standard. Sometimes, differences in experience appear among medical researchers participating in an evaluation and leveling the experience does not seem feasible. In such cases, the evaluators choose alternative standards or create custom questionnaires (Paper 2).

This thesis has also shown that the expectations of using standards as an assurance to quality in eHealth evaluations fail to recognize the problematics of organizational complexity. Barriers to standards' applicability emerge in a multiparty evaluation, when researchers from different disciplines and contexts have to reach consensus regarding the common standards to be used during the evaluation. In this activity, the parties bring their goals, resources, and capabilities, and translate the standards to fit own circumstances (Paper 1 and Paper 2). Translation (Latour, 1987; Røvik, 2007; Czarniawska and Sevón, 2011) can result in alterations of the original content of a standard. Hence, variability in circumstances of the research teams can impact the applicability of standards and lead to various degrees of adherence to their content (Paper 2). Adherence problems have also been observed in the application of standardized clinical guidelines to real care processes by healthcare professionals (Burstin et al., 1999; Grilli and Lomas, 1994). However, it should be noted that translation should not be always seen in a negative light. The quality of evaluation can also be diminished by applying standards in full adherence that do not fit with the study context and no better alternatives are available. The findings of Paper 2 support previous research, claiming that evaluations can be challenging owing to a complex social reality (Stufflebeam, 2001; Bates and Wright, 2009) that requires

alignment of goals, resources, and capabilities (Greenhalgh and Russel, 2010).

5.2 *RQ2 – How can value of an eHealth solution be conceptualized and measured?*

Most summative eHealth evaluation frameworks do not explicitly focus on identifying the value of an eHealth solution. Implicitly, however, they describe the same thing: capturing something's merit or worth (Mark et al., 2000) and investigating whether care supported by eHealth is better than the traditional mode of care (Kolasa and Kozinski, 2020). In the studies that explicitly address value (Bergmo, 2015; Talboom-Kamp et al., 2016; Runz-Jørgensen et al., 2017; Ariens et al., 2017), the conceptualization of value of eHealth has been vague and an array of concepts have been referred to as value (Paper 4). A lack of methodological base for capturing value of eHealth can arise from the fact that evaluating eHealth needs addressing medical, IT, social, and business fields (Eysenbach, 2001; Kaplan and Shaw, 2004) where problems arise from differences in ontology, research traditions and established research methods.

This problematization of capturing value through a summative evaluation shows that there is a need to create specific models for capturing value in the dynamic and multi-stakeholder context of eHealth. Paper 4 offers a step towards a better conceptualization of value by providing conceptual categories for otherwise disparate evaluation domains (as in, e.g., Lampe et al., 2009; Kidholm et al., 2012) (see Figure 5). Paper 4 defines value of eHealth as a holistic view of the created monetary and nonmonetary benefits of eHealth that require monetary and nonmonetary sacrifices in a particular context. The proposed model for evaluating value of an eHealth service (hereinafter, the model) consists of four domains, namely benefits (monetary and non-monetary) and sacrifices (monetary and non-monetary).

Context

Benefits	Sacrifices
Monetary	Monetary
Non-monetary	Non-monetary

Figure 5. Structure of the value assessment of an eHealth intervention

The model differs from previous evaluation frameworks (e.g., Lampe et al., 2009; NICE, 2019; Kolasa and Kozinski, 2020) through its different conceptual portrayal of traditional evaluation domains (such as clinical performance, organizational impact, technical usability, and economic impact) through a lens of benefits and sacrifices that can be assessed for different

stakeholders (Paper 4). For instance, previous evaluation frameworks have included evaluation domains such as cost effectiveness and organizational changes, but these were presented as separate elements, lacking logic regarding how these are inter-connected (like in, e.g., Kidholm et al., 2012; Andargoli et al., 2017), a problem identified by previous research (Kolasa and Kozinski, 2020). In addition, assessing organizational changes can be problematic when there is duality in conceptual understanding, as organizational changes can also be viewed as positives, negatives, or a combination of both. Novelty in the proposed model (Figure 5) manifests through (1) the conceptual "tying" of the evaluation themes together and not showing them as a mere collection of separate themes, (2) emphasized sacrifice that draws the evaluator's attention that the downsides (both financial and non-financial) of introducing and using an eHealth service should not be overlooked and should be accounted for. In other words, the benefits will come at a certain expense, and sacrifices will be required when deploying the service.

The usefulness of this model manifests when the four domains are contrasted against each other. For example, staff workload (a non-monetary sacrifice) can be considered in light of achieved health outcomes and patient experience (non-monetary benefits). Such a comparison provides a more holistic and fair view on the value of an eHealth solution. In summary, this model can be particularly useful in summative evaluations that aim to produce evidence for decision making regarding deployment of an eHealth service within a healthcare system. The model could be used by various stakeholders by populating the model with the variables or themes relevant to them. However, it should be noted that patients usually do not have access to monetary data such as treatment costs, hourly rates of staff, or prices of technology. Therefore, the usefulness of the model for patients as stakeholders in the value assessment is limited. Adopting the proposed conceptualization of value of eHealth and the model for its assessment can increase methodological uniformity in the eHealth research (useful in cross-country evaluations and systematic reviews) and contribute to enhancing quality of eHealth studies.

5.3 *RQ3 – How is evidence from an eHealth trial used to support improvements in healthcare?*

Building decisions on hard evidence from rigorous evaluations stems from the evidence-based medicine, a pharmacological approach that underlies eHealth evaluation methodologies (Evidence-Based Medicine Working Group, 1992). Decisions of prescribing drugs are based on national guidelines that are built on the best available research evidence (summative evaluations). When this evidence is not available, decisions need to be informed by professional judgement (Evidence-Based Medicine Working Group, 1992). However, research described in this thesis showed that this approach is too simplistic and crude for the eHealth context (Paper 5). Various stakeholders use evidence created by a summative evaluation for learning purposes and in scientific publishing, applying for research funding, and supporting regional policies

(Paper 5). Interestingly, evaluation evidence (including clinical outcomes) sometimes seems to "weigh less" in making care improvement decisions than the opinions of healthcare professionals. These opinions can include insights regarding usability, safety, and user satisfaction that are formed while trying out the solution. These opinions can lead to decisions to adopt the technology even without a complete summative evaluation (Paper 5). Hence, the case study analyzed in this thesis showed that the role of evidence from a summative evaluation seems to be smaller than traditionally assumed. This gap raises considerations regarding the usefulness of summative evaluations in improving care.

One reason for the limited usefulness of summative evaluations could stem from the fact that eHealth is a crossroad between medical, IT, social, and business areas (Eysenbach, 2001; Shaw et al., 2017), and the fact that combination of decision-making rules applicable in these fields (Glasgow, 2007; Rigby et al., 2018) might not directly correspond with the rules of pharmacology or health technology assessment. Instead, improvement decisions involving eHealth are built on the value associated with the technology. Healthcare professionals create understandings of the value while trying out the technology (Barlow, 2016) and by navigating the social dynamics among multiple involved stakeholders (Zuiderent-Jerak and Bruun Jensen, 2007) involved in eHealth research, which culminates in a summative evaluation. Trust by healthcare professionals in professional judgement for making improvement decisions involving eHealth (especially for non-invasive technologies) might emerge from an understanding of own agency, or "if we like it, we can make it work", and the intentionality embedded in research based on interventions (Zuiderent-Jerak and Bruun Jensen, 2007). Since healthcare professionals partially construct and control the transformational processes of eHealth research, it is important that an evaluation design allows them to form experiences and opinions with the technology (Paper 5). Professional judgement might also be preferred in making practice improvement decisions due to the inherent capabilities of humans in capturing contextual sensitivities. In contrast, evidence from summative evaluations might be less valuable due to its simplified view on the reality (simplification through standardization) (Sager and Zuiderent-Jerak, 2020).

In addition, it seems that the role of healthcare professionals involved in eHealth interventions can extend beyond delivering the intervention to the patients, being the data collectors or interview subjects regarding usability, safety, or satisfaction with the technology (e.g., Runz-Jørgensen et al., 2017; Ariens et al., 2017). Research described in this thesis shows that, through using the conceptual use of evidence, they might take on the role of advocates for or against the solution's deployment into practice and they can ultimately influence to which improvements investments are allocated (Paper 5). Therefore, it is important to strengthen the conceptual use of evidence to create actual changes in healthcare practice, and the research funding bodies should regard the conceptual use of evidence as a success of the project. The conceptual use of evidence can be supported by promoting active participation of healthcare professionals in eHealth interventions and stimulating reflections upon the experience gained.

Finally, limited usefulness of evidence from a summative eHealth evaluation has shown that an understanding of quality in evaluations should not be limited to scientific rigor, such as adherence to standards, use of comparison group, application of randomized procedures, systematic sample size calculation, data collection at baseline and end-line, and independence of evaluators (Bates and Wright, 2009; Lilford et al., 2009; Mookherji et al., 2015). A quality evaluation should bring value to its users in order to make the effort worthwhile, and methodologies of eHealth evaluations should be oriented towards this value creation.

5.4 *Theoretical contributions*

This thesis has extended research on eHealth evaluations by focusing on the quality problems in evaluations. These efforts revealed tensions between evaluation quality criteria 'contextual sensitivity' and 'adherence to standards' and showed that it can be difficult to meet both of these criteria. By drawing on the translation theory (Latour, 1987; Røvik, 2007; Czarniawska and Sevón, 2011), translation strategies (Røvik, 2007), the process framework of the development of cooperative inter-organizational relationship (Ring and Van de Ven, 1994), the thesis identified barriers determining trade-offs between contextual circumstances and the standards' content that evaluators make. These barriers explain why published evaluations lack methodological uniformity (Greenhalgh et al., 2004; Ammenwerth, 2004; Proudfoot et al., 2011; Janssen et al., 2013a) and can be a sound basis for taking action in improving the methodologies for eHealth evaluation. Further, the thesis extends knowledge on the quality criteria for eHealth evaluations identified by previous research (Catwell and Sheikh, 2009; Greenhalgh and Russel, 2010; Nilsen et al., 2012; Mookherji et al., 2015; White et al., 2016; Pham et al., 2016; Andargoli et al., 2017; Dick et al., 2020) and by adding two new criteria "Capturing value of an eHealth solution" and "Involving healthcare professionals in the intervention and its evaluation".

Building on the established value conceptualizations in service management (Ramirez, 1999; Grönroos, 2008; Gummerus, 2013), the present thesis also contributes with a conceptualization of a 'value' concept in the context of eHealth offering a unified conceptual definition that could serve as a foundation for methodological developments in eHealth evaluation. The model for assessing value in eHealth interventions, another contribution of this thesis (Figure 5), can be a beneficial basis towards these efforts.

Finally, the thesis draws attention on the importance of 'validating' worthiness of an evaluation with the potential users of the generated evidence. By employing theory of evaluation use (Alkin and Taut, 2002; Leviton and Hughes, 2016; Alkin and King, 2017), the thesis contributes insights into what factors reduce the chances that evidence from evaluations would be used to provide valuable feedback to research of eHealth evaluations.

5.5 *Practical contributions*

The findings in this thesis lead to several practical propositions for eHealth evaluation practitioners. Some of the propositions are also applicable to the professional organizations developing eHealth evaluation standards. A list of criteria for quality in eHealth evaluations has been summarized and extended in this thesis (Table 4). Evaluation practitioners could benefit from using these criteria in planning summative evaluations. The professional organizations (for example, the ones responsible for standardization) could aim to support evaluators with better guidance aligned with these criteria. This thesis also confirms worthiness in assessing standards as recommended by previous research (Mookherji et al., 2015; Cowie et al., 2016). Aligned with these recommendations, the thesis has identified gaps in different types of eHealth evaluation standards and uncovered the barriers that lead to evaluators struggling to apply and adhere to these standards in eHealth evaluations. Addressing and solving these identified hindrances might increase scientific rigor and could lead to a higher methodological uniformity (useful in cross-country evaluations and systematic reviews) among eHealth evaluation studies (Dick et al., 2020).

When planning a summative eHealth evaluation, evaluators are recommended to apply a model for eHealth value assessment as a framework (Figure 5). Monetary and non-monetary benefits and sacrifices, which are the domains of the framework, should be populated with the appropriate evaluation themes in accordance with the contract of the study (for external evaluators) and the research questions. By applying this model, assessments of eHealth interventions could deliver more holistic evidence regarding the merits and worth (Mark et al., 2000) of an eHealth solution in a particular context.

By building on the previously identified gap of lacking processual studies in healthcare innovations (Greenhalgh et al., 2004) and the model for continuous improvement (Deming, 1956), the present thesis has shown that it is worthwhile studying eHealth evaluations throughout their life cycles (from planning an evaluation to studying the actual use of evaluation results) in order to identify various problems in eHealth evaluations and how to solve them. This approach is suggested to use in future studies aimed at improving quality of eHealth evaluations.

5.6 *Reflections on actions to improve quality in eHealth evaluations*

This thesis has demonstrated that studying evaluations longitudinally and throughout their lifecycle can reveal an in-depth and nuanced view on the challenges and dynamics of evaluations. Such an approach is in line with calls for more process studies on innovation in healthcare (Greenhalgh et al., 2004). Therefore, eHealth evaluation studies, especially those oriented towards improving evaluation methodology and practice, could benefit from applying a Plan-Do-Study-Act cycle (PDSA), a model for improvement (Deming, 1952) (Figure 6).

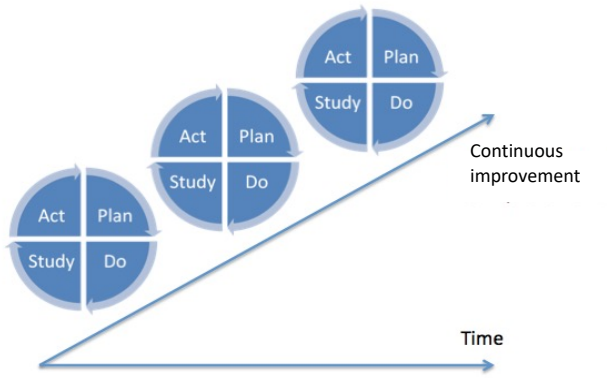

Figure 6. PDSA cycle

The model should be applied iteratively in order to identify problems, solve them, and study their impact. In the context of eHealth, it could be applicable to eHealth evaluation by conducting and studying one or multiple evaluations over time, identifying gaps, studying for possible solutions, and improving methodologies or practice (for example, the next evaluation). In accordance with the purpose of this thesis, the longitudinal approach of studying an evaluation also resembled the PDSA model, since the eHealth evaluation was studied throughout its planning (Plan), conducting (Do), and reflecting on methodology and evidence (Study) phases. The thesis ends with actionable insights (Act) that can be considered during the improvement efforts to increase quality in eHealth evaluations (see Figure 7).

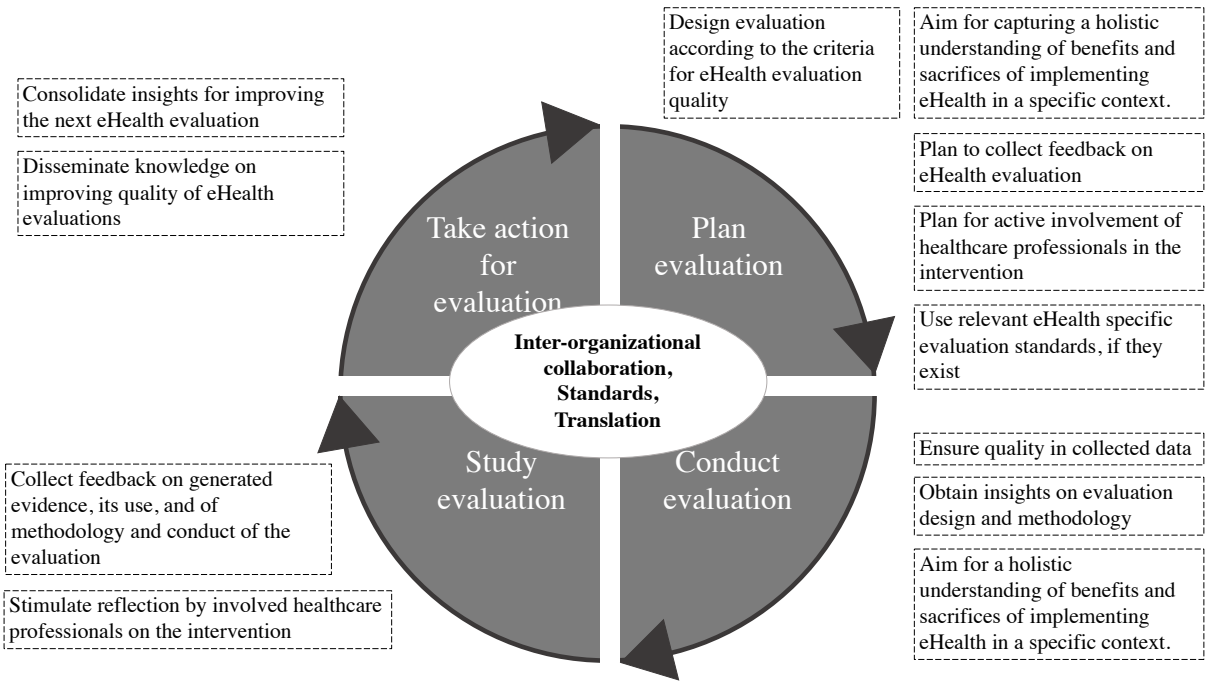

Figure 7. A model for improving quality in eHealth evaluations

Plan evaluation

The plan phase of an eHealth evaluation needs to be informed by the planning standards (e.g., Nykänen et al., 2011; Cusack et al., 2009; European Commission, 2016), evaluation frameworks (e.g., Kidholm et al., 2012) and criteria for evaluation quality, as outlined in Table 6 (Ammenwerth, 2004; Greenhalgh et al., 2004; Proudfoot et al., 2011; Janssen et al., 2013a; Mookherji et al., 2015; Cowie et al., 2016; Paper 1; Paper 2; Paper 5). When possible and relevant, eHealth-specific evaluation standards should be used (Paper 1, Paper 2). When multiple organizations collaborate on the evaluation, it is important to consider the aims and agendas of these stakeholders already in the planning phase (for example, it might affect which standards should be used). The methodology of evaluation and generated evidence should help these organizations in fulfilling the agendas.

Furthermore, it is advisable to plan to capture value of an eHealth solution in a specific context. This could be done by considering which benefits and sacrifices of implementing an eHealth solution should be measured and assessed. It is also recommended to plan to involve healthcare professionals in the intervention and its evaluation activities, since it promotes the use of evidence and adoption of the eHealth solution in practice (Paper 5). The plan should present the scope and should function as a roadmap of how the evaluation shall be conducted. To improve evaluations, regardless of the context and overall aims of evaluation, it is suggested to also plan for how feedback on evaluation will be collected. Planning and orientation for learning could help in terms of taking deliberate actions to study and improve evaluations, which can lead to improvements in methodologies and competences.

Conduct evaluation

During the conduct phase of an eHealth evaluation, it is essential to focus on quality in order to ensure quality and completeness in data (Paper 4; Paper 5). This helps to achieve the aims of evaluation, enables determination of value of an eHealth solution (Paper 4), and supports the use of created evidence (Paper 4). Issues arising from the conducting phase should inform the studying phase and should be an input (feedback) to the planning phase.

Study evaluation

The study phase provides insights into the use of evidence and the evaluation methodology and its conduct through reflective activities or interviews with the beneficiaries of evidence (Paper 5). Such efforts not only provide feedback and help improve planning of the next evaluations, but also build new knowledge by learning about the challenges during evaluation, usefulness of the generated evidence, and barriers to evidence use as experienced by different stakeholders (Van Aartsengel and Kurtoglu, 2013; Paper 5). Furthermore, it is recommended to stimulate reflective activities by the healthcare professionals involved in the intervention, since this allows for the sharing of experiences and opinions formed and promotes the use of evidence and adoption of the eHealth solution in practice (Paper 5).

Take action for evaluation

During the acting phase it is decided what should be done during planning of the next evaluation depending on the outcome of the analysis in the study phase. Necessary modifications to eHealth evaluation methodology and organization should be determined. If possible and relevant, the findings from the conducting and studying phases should be published in order to further disseminate knowledge and achieve changes in the field of the eHealth evaluation.

6 CONCLUSIONS AND FUTURE RESEARCH

This chapter presents the conclusions, limitations, and suggestions for future research.

6.1 *Conclusions*

The purpose of this thesis was to study how summative eHealth evaluations can be improved to support the determination of eHealth value in a specific context, as well as the use of evidence produced during eHealth interventions. The findings are built on four studies performed throughout a lifecycle of a single case of a summative eHealth evaluation. Building on the findings of the studies, the purpose of this thesis has been addressed by reflecting on the evaluation quality criteria.

The key contribution of this thesis is twofold. First, the extended view on eHealth evaluation quality is presented by providing deeper insight on the problems in the existing quality criteria and by introducing a new criterion for quality. The thesis reveals that meeting some of the criteria is not always practical and that evaluators might make trade-offs between the criteria. The findings point to a need to improve methodologies for eHealth evaluations by providing better guidance to evaluators and validating evaluation standards in different locations.

Second, the value of an eHealth solution requires a unified definition and a methodological framework for assessment. By viewing value as a monetary and nonmonetary benefits and sacrifices in a particular context, a more complete evidence on the value an eHealth solution can be generated. The thesis also suggests that value is relative depending on the context and generalizability of evidence on value is limited. The proposed model for value assessment is the first step towards value-oriented evaluation methodologies in the context of eHealth.

6.2 *Limitations*

The research described in this thesis has several limitations. First, all of the studies were based on a single case multi-national and interdisciplinary research project and the findings were contextualized in accordance with the set-up of the project and the collaborating partners. A multiple-case study or different set-ups might have provided a different or wider spectrum of evaluation practices and quality issues. In addition, the case studied involved commercially available eHealth solutions, which means their maturity was satisfactory. The thesis has not addressed evaluations of early-stage technologies, which could have produced different results in terms of evaluation quality and the use of evidence after the intervention. Hence, the generalizability of the results and conclusions can be troublesome. However, a single case can contribute to theoretical generalization; that is, the results and conclusions can be used to further

develop theory on eHealth evaluation. Further, specific additional limitations of every study are discussed.

In Paper 1, in addition to the eHealth evaluation planning guidelines that were included in the study, there might be additional ones that, if included, might have provided a more complete set of recommendations for improving the guidelines. Also, there might be a portion of e-mails and other internal communication documents that are not available to the authors affecting the data set in this study. For example, informal conversations that occurred outside the formal meetings or e-mail communication were not included in the data set.

In Paper 2, including more standards in the study could have possibly provided more factors affecting changes in application of evaluation standards. Additionally, the main data were collected from e-mail correspondence between the partners in a research consortium. Other means of data collection were not explored in this study. In addition, informal conversations that took place outside the formal meetings or e-mail communication were not included in the data set. Furthermore, organizational hierarchies between the stakeholders were not considered during data analysis, meaning that opinions by some organizational roles could be underrepresented.

In Paper 3, the main limitation of the study was the follow-up period of six months, which affected the observed outcomes in patient health and quality of life. A longer period could have possibly provided more significant results. Another limitation was related to the outcomes that could not be turned into monetary values (both for costs and benefits). In such cases, qualitative discussion is provided. Lastly, there was a lack of preventable falls data since the data were not collected in the study and had to be obtained from the literature. However, the literature used is based on the same OTAGO program in other studies.

In Paper 4, including more countries for comparing results of summative evaluations between countries could have provided a more extensive list of the identified contextual factors affecting the value of the eHealth solution. Studying two countries in depth demanded time and effort and it was not feasible to include more countries. Also, the study was constrained by a six-month follow-up time for the patients, which affected clinical effectiveness outcomes. A longer follow-up time could produce long-term effects and allow for additional contextual factors to appear, which could possibly impact the conceptualization of value.

In Paper 5, exploration of the use of evidence from the eHealth intervention was based on a specific research project setup and the interviewee sample was limited to the partners of the research consortium, since they had deep knowledge of the project and its results and were in a favorable position to use the evidence obtained. Other means of data collection were not explored. For example, information from the informal conversations among the stakeholders might have provided a deeper understanding of the motivations behind the use of evidence. Furthermore, organizational hierarchies between the stakeholders were not considered during

data analysis, meaning that opinions by some organizational roles could be underrepresented in the present study. Evidence use by additional stakeholders outside the DECI consortium were not explored such as the funding agency, industry, or government. Similarly, evidence produced in other settings and study designs could provide a different view of the use of evidence. Furthermore, the study captured the situation eight months after the project was finished. However, the use of evidence might be more extensive in later stages due to the so-called "gestation period" (Feinstein, 2002).

6.3 *Future research*

Future research should aim to contribute to creating the eHealth-specific standards. A good starting point could be to further conceptualize the hurdles of standardization and ways of producing knowledge in eHealth evaluations. Such efforts could also include analyzing problems in applying a range of standards from different fields to the context of eHealth, or in different research set-ups. Another avenue for future research could be exploring the "voice" of eHealth evaluation practitioners as potential adopters of standards to determine effective mechanisms that can enhance the uptake of standards in eHealth evaluation practice.

Future studies could explore how to understand and capture the locally meaningful value of an eHealth solution through summative evaluations by considering that it is embedded in a broader framework of services. It could also be worth further investigating factors determining differences in eHealth value manifesting in different geographical contexts, stakeholder networks, and research set-ups. Such knowledge could provide better insights on the generalizability of eHealth research conducted in a particular context, and it could feed back into evaluation methodologies and help to further conceptualize value.

The thesis provides insights regarding the limited usefulness of summative evaluations in supporting eHealth adoption decisions. Future research could further explore what evidence is meaningful to different stakeholders in their decisions, and what methods of producing evidence 'count'. In addition, it could be worthwhile investigating the actual use of evidence generated through summative evaluations in different time frames – during data collection, after evaluation report, after several months and longer – to increase the chances of capturing full benefits of the produced evidence.

Finally, future studies should monitor progress in eHealth evaluation quality and in the way in which studies are designed, in order to track changes in evaluation quality problems. Such attempts might require more in-depth case studies.

REFERENCES

Alla, K., Oprescu, F., Hall, W. D., Whiteford, H. A., Head, B. W., & Meurk, C. S. (2018). Can automated content analysis be used to assess and improve the use of evidence in mental health policy? A systematic review. *Systematic Reviews*, *7*(1), 1–16.

Allen, R. H., & Sriram, R. D. (2000). The role of standards in innovation. *Technological Forecasting and Social Change*, *64*(2–3), 171–181.

Alkin, M. C. (ed.) (2004). *Evaluation roots*. Thousand Oaks: SAGE.

Alkin M. C., & Taut S. M. (2002). Unbundling evaluation use. *Studies in Educational Evaluation, 29*(1), 1–12.

Alkin M. C., & King J. A. (2017). Definitions of Evaluation Use and Misuse, Evaluation Influence, and Factors Affecting Use. *American Journal of Evaluation*, *38*(3):434–450.

Ammenwerth, E., Brender, J., Nykänen, P., Prokosch, H. U., Rigby, M., & Talmon, J. (2004). Visions and strategies to improve evaluation of health information systems: Reflections and lessons based on the HIS-EVAL workshop in Innsbruck. *International Journal of Medical Informatics*, *73*(6), 479–491.

Ammenwerth, E., & Rigby, M. (eds.) (2016). *Evidence-Based Health Informatics: Promoting Safety and Efficiency Through Scientific Methods and Ethical Policy*. Amsterdam, The Netherlands: IOS Press.

Andargoli, A. E., Scheepers, H., Rajendran, D., & Sohal, A. (2017). Health information systems evaluation frameworks: A systematic review. *International Journal of Medical Informatics*, *97*, 195–209.

Andersen, H., Røvik, K. A., & Ingebrigtsen, T. (2014). Lean thinking in hospitals: Is there a cure for the absence of evidence? A systematic review of reviews. *BMJ Open*, *4*(1), e003873.

Andreassen, H. K., Kjekshus, L. E., & Tjora, A. (2015). Survival of the project: a case study of ICT innovation in health care. *Social Science & Medicine*, *132*, 62–69.

Ariens, L. F., Schussler-Raymakers, F. M., Frima, C., Flinterman, A., Hamminga, E., Arents, B. W., Bruijnzeel-Koomen, C. A. F. M., de Bruin-Weller, M. S., & van Os-Medendorp, H. (2017). Barriers and facilitators to eHealth use in daily practice: perspectives of patients and professionals in dermatology. *Journal of Medical Internet Research*, *19*(9), e300.

Aqil, A., Lippeveld, T., & Hozumi, D. (2009). PRISM framework: a paradigm shift for designing, strengthening and evaluating routine health information systems. *Health Policy and Planning*, *24*(3), 217–228.

Barlow, J. (2016). *Managing innovation in healthcare*. World Scientific Publishing Company.

Batalden, M., Batalden, P., Margolis, P., Seid, M., Armstrong, G., Opipari-Arrigan, L., & Hartung, H. (2016). Coproduction of healthcare service. *BMJ Quality & Safety*, *25*(7), 509–517.

Bates, D. W., & Wright, A. (2009). Evaluating eHealth: undertaking robust international cross-cultural eHealth research. *PLoS Medicine*, *6*(9), e1000105.

Bergmo, T. S. (2015). How to measure costs and benefits of eHealth interventions: an overview of methods and frameworks. *Journal of Medical Internet research*, *17*(11), e254.

Berwick, D. M., Nolan, T. W., & Whittington, J. (2008). The triple aim: care, health, and cost. *Health Affairs*, *27*(3), 759–769.

Bickman, L. (1987). The functions of program theory. *New Directions for Program Evaluation*, (33), 5–18.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101.

Brunsson, N., & Jacobsson, B. (2000). *A world of standards*. Oxford, England: Oxford University Press.

Bryman, A., & Bell, E. (2011). *Business research methods*. Oxford, England: Oxford University Press.

Burstin, H. R., Conn, A., Setnik, G., Rucker, D. W., Cleary, P. D., O'Neil, A. C., Orav, E. J., Sox, C. M., Brennan, T. A. & Harvard Emergency Department Quality Study Investigators. (1999). Benchmarking and quality improvement: the Harvard emergency department quality study. *The American Journal of Medicine*, *107*(5), 437–449.

Catwell, L., & Sheikh, A. (2009). Evaluating eHealth interventions: the need for continuous systemic evaluation. *PLoS Med*, *6*(8), e1000126.

Chouinard, J. A., & Cousins, J. B. (2009). A review and synthesis of current research on cross-cultural evaluation. *American Journal of Evaluation*, *30*(4), 457–494.

Cohen, G., Schroeder, J., Newson, R., King, L., Rychetnik, L., Milat, A. J., Bauman, A. E., Redman, S., & Chapman, S. (2015). Does health intervention research have real world policy and practice impacts: testing a new impact assessment tool. *Health Research Policy and Systems*, *13*(1), pp.1–12.

Collins, L. M., Murphy, S. A., Nair, V. N., & Strecher, V. J. (2005). A strategy for optimizing and evaluating behavioral interventions. *Annals of Behavioral Medicine*, *30*(1), 65–73.

Collins, L. M., Murphy, S. A., & Strecher, V. (2007). The multiphase optimization strategy (MOST) and the sequential multiple assignment randomized trial (SMART): new methods for more potent eHealth interventions. *American Journal of Preventive Medicine*, *32*(5), S112–S118.

Cowie, M. R., Bax, J., Bruining, N., Cleland, J. G., Koehler, F., Malik, M., Pinto, F., Van der Velde, E. & Vardas, P. (2016). e-Health: a position statement of the European Society of Cardiology. *European Heart Journal*, *37*(1), 63–66.

Cusack C. M., Byrne C., Hook J. M., McGowan J., Poon E. G., Zafar A. (2009). Health Information Technology Evaluation Toolkit: 2009 Update (Prepared for the AHRQ National Resource Center for Health Information Technology under Contract No. 290-04-0016.) AHRQ Publication No. 09-0083-EF. Rockville, MD: Agency for Healthcare Research and Quality.

Czarniawska, B., & Sevón, G. (Eds.). (2011). *Translating Organizational Change* (Vol. 56). Berlin: de Gruyter.

De Keizer, N. F., & Ammenwerth, E. (2008). The quality of evidence in health informatics: How did the quality of healthcare IT evaluation publications develop from 1982 to 2005? *International Journal of Medical Informatics*, *77*(1), 41–49.

Denzin, N. K. (1970). *The Research Act in Sociology*. Chicago: Aldine.

Dick, S., O'Connor, Y., Thompson, M. J., O'Donoghue, J., Hardy, V., Wu, T. S. J., O'Sullivan, T., Chirambo, G. B., & Heavin, C. (2020). Considerations for Improved Mobile Health Evaluation: Retrospective Qualitative Investigation. *JMIR mHealth and uHealth*, *8*(1), e12424.

Deming, W. E. (1952). *Elementary principles of the statistical control of quality: a series of lectures*. Nippon Kegaku Gijutsu Remmei.

Diener, E., & Crandall, R. (1978). *Ethics in social and behavioral research*. University of Chicago Press.

Diviani, N. (2014). Validation of an Italian Version of the EHealth Literacy Scale (I-EHEALS). In *Medicine 2.0 Conference*. Toronto, Canada: JMIR Publications Inc.

Djulbegovic, B., & Guyatt, G. H. (2017). Progress in evidence-based medicine: a quarter century on. *The Lancet*, *390*(10092), 415–423.

Doherty, G., Coyle, D., & Matthews, M. (2010). Design and evaluation guidelines for mental health technologies. *Interacting with Computers*, *22*(4), 243–252.

Donaldson, S. I., & Christie, C. A. (2006). Emerging career opportunities in the transdiscipline of evaluation science. In S. I. Donaldson, D. E. Berger, & K. Pezdek (eds) *Applied Psychology: New frontiers and rewarding careers*, 243–260. Mahwah, NJ: Erlbaum.

Donaldson, S. I., & Lipsey, M. W. (2006). Roles for theory in evaluation practice. *Handbook of Evaluation*. California, Thousand Oaks: SAGE.

Elf, M., Flink, M., Nilsson, M., Tistad, M., von Koch, L., & Ytterberg, C. (2017). The case of value-based healthcare for people living with complex long-term conditions. *BMC Health Services Research*, *17*(1), 1–6.

Erikson, K. T. (1967). A comment on disguised observation in sociology. *Social Problems*, 14, 366–73.

Erlandson, D. A., Harris, E. L., Skipper, B. L., Allen, S. D. (1993). *Doing Naturalistic Inquiry. A Guide to Methods*. Newbury Park: SAGE.

European Commission's Directorate General for Communications Networks, Content & Technology (DG CONNECT), https://ec.europa.eu/digitalcsinglec market/en/news/reportceuccountriescplancspendc954cmillioncehealthc2014c2020, accessed 28 June 2016.

European Commission. MEDDEV 2.7/1 – Revision 4. Guideline, 2016. *Available:* https://www.google.co.uk/search?sxsrf=ACYBGNS4e49fPap_w2zhkg7WX_fYwFEXEg%3A1581441138816&source=hp&ei=cuBCXuagL8eoa7OOq9AI&q=MEDDEV+2.7%2F1+revision+4&oq=MEDDEV+2.7%2F1+revision+4&gs_l=psy-ab.3.0j0i22i30l9.2869.5700.6015.1.0.0.67.660.11.0.2j1. [Accessed 3 May 2020]

EuroQol Group. EuroQol – A new facility for the measurement of health-related quality of life. (1990). *Health Policy, 16*(3), 199.

Evans, D. (2003). Hierarchy of evidence: a framework for ranking evidence evaluating healthcare interventions. *Journal of Clinical Nursing*, *12*(1), 77–84.

Evidence-Based Medicine Working Group. (1992). Evidence-based medicine. A new approach to teaching the practice of medicine. *JAMA, 268*(17):2420–2425

Eysenbach, G. (2001). What is eHealth? *Journal of Medical Internet Research*, *3*(2):e20.

Farrell, J., & Saloner, G. (1985). Standardization, compatibility, and innovation. *the RAND Journal of Economics*, 70–83.

Feinstein, O. N. (2002). Use of Evaluations and the Evaluation of their Use. *Evaluation*, *8*(4), 433–439.

Flick, U. (2014). *An Introduction to Qualitative Research*. London, UK: SAGE.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*(3), 189–198.

Gehman, J., Trevino, L. K., & Garud, R. (2013). Values work: A process study of the emergence and performance of organizational values practices. *Academy of Management Journal*, *56*(1), 84–112.

Glasgow, R. E. (2007). eHealth evaluation and dissemination research. *American Journal of Preventive Medicine*, *32*(5), S119–S126.

Glick, W. H., Huber, G. P., Miller, C. C., Doty, D. H., & Sutcliffe, K. M. (1990). Studying changes in organizational design and effectiveness: Retrospective event histories and periodic assessments. *Organization Science*, *1*(3), 293–312.

Greenhalgh, T., Robert, G., Macfarlane, F., Bate, P., & Kyriakidou, O. (2004). Diffusion of innovations in service organizations: systematic review and recommendations. *The Milbank Quarterly*, *82*(4), 581–629.

Greenhalgh, T., Wherton, J., Papoutsi, C., Lynch, J., Hughes, G., A'Court, C., Hinder, S., Fahy, N., Procter, R., & Shaw, S. (2017). Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and

sustainability of health and care technologies. *Journal of Medical Internet Research*, *19*(11).
Greenhalgh, T., & Russell, J. (2010). Why do evaluations of eHealth programs fail? An alternative set of guiding principles. *PLoS Medicine*, *7*(11), e1000360.
Greenhalgh, T., Howick, J., & Maskrey, N. (2014). Evidence based medicine: a movement in crisis? *BMJ*, *348*.
Grilli, R., & Lomas, J. (1994). Evaluating the message: the relationship between compliance rate and the subject of a practice guideline. *Medical Care*, 202–213.
Grönroos, C. (2008). Service logic revisited: who creates value? And who co-creates? *European Business Review*, *20*(4), 298–314.
Guba, E. G., & Lincoln, Y. S. (1994). Competing paradigms in qualitative research. *Handbook of Qualitative Research*, *2*(163–194), 105.
Guba, E. G., & Lincoln, Y. S. (1989). *Fourth generation evaluation*. Newbury Park, CA: SAGE.
Gummerus, J. (2013). Value creation processes and value outcomes in marketing theory: strangers or siblings? *Marketing Theory*, *13*(1), 19–46.
Gurman, T. A., Rubin, S. E., & Roess, A. A. (2012). Effectiveness of mHealth behavior change communication interventions in developing countries: a systematic review of the literature. *Journal of Health Communication*, *17*(sup1), 82–104.
Halldorsson, A., & Aastrup, J. (2003). Quality criteria for qualitative inquiries in logistics. *European Journal of Operational Research*, *144*(2), 321–332.
Homan, R. (1991). *The ethics of social research*. Addison-Wesley Longman Limited.
Hellström, A., Lifvergren, S., & Quist, J. (2010). Process management in healthcare: investigating why it's easier said than done. *Journal of Manufacturing Technology Management*, *21*(4), 499–511.
International Medical Informatics Association, IMIA Yearbook of Medical Informatics 2013: Evidence-Based Health Informatics, Stuttgart: Schattauer, 2013.
Janssen, R., Hettinga, M., Visser, S., Menko, R., Prins, H., Krediet, I., Haaker, T. & Bodenstaff, L. (2013a). Innovation routes and evidence guidelines for eHealth Small and Medium-sized Enterprises. *International Journal on Advances in Life Sciences*, *5(3&4),* 188–203.
Kaplan, B. (2001). Evaluating informatics applications – some alternative approaches: theory, social interactionism, and call for methodological pluralism. *International Journal of Medical Informatics*, *64*(1), 39–56.
Kaplan, B., & Shaw, N. T. (2004). Future directions in evaluation research: people, organizational, and social issues. *Methods of Information in Medicine-Methodik der Information in der Medizin*, *43*(3), 215–231.
Kazanjian, A., & Green, C. J. (2002). Beyond effectiveness: the evaluation of information systems using a comprehensive health technology assessment framework. *Computers in Biology and Medicine*, *32*(3), 165–177.
Kelly, L., Ziebland, S., & Jenkinson, C. (2015). Measuring the effects of online health information: scale validation for the e-health impact questionnaire. *Patient Education and Counseling*, *98*(11), 1418–1424.
Khoja, S., Durrani, H., Scott, R. E., Sajwani, A., & Piryani, U. (2013). Conceptual framework for development of comprehensive e-health evaluation tool. *Telemedicine and e-Health*, *19*(1), 48–53.
Kidholm, K., Ekeland, A. G., Jensen, L. K., Rasmussen, J., Pedersen, C. D., Bowes, A., Flottorp, S.A., & Bech, M. (2012). A model for assessment of telemedicine applications: mast. *International Journal of Technology Assessment in Health Care*, *28*(1), 44–51.

Kip, H., Oberschmidt, K., & Bierbooms, J. J. (2020). Ehealth technology in forensic mental healthcare: recommendations for achieving benefits and overcoming barriers. *International Journal of Forensic Mental Health*, *20*(1), 31–47.

Klasnja, P., Hekler, E. B., Shiffman, S., Boruvka, A., Almirall, D., Tewari, A., & Murphy, S. A. (2015). Microrandomized trials: An experimental design for developing just-in-time adaptive interventions. *Health Psychology*, *34*(S), 1220.

Klein, H. K., & Myers, M. D. (1999). A set of principles for conducting and evaluating interpretive field studies in information systems. *MIS Quarterly*, 67–93.

Kolasa, K., & Kozinski, G. (2020). How to Value Digital Health Interventions? A Systematic Literature Review. *International Journal of Environmental Research and Public Health*, *17*(6), 2119.

Koppel, R. (2013). Is healthcare information technology based on evidence? *Yearbook of Medical Informatics*, *22*(01), 7–12.

Kvale, S. (2007). *Doing Interviews*. London: Sage.

Lampe, K., Mäkelä, M., Garrido, M. V., Anttila, H., Autti-Rämö, I., Hicks, N. J., Hofmann, B., Koivisto, J., Kunz, R., Kärki, P. & Malmivaara, A. (2009). The HTA core model: a novel method for producing and reporting health technology assessments. *International Journal of Technology Assessment in Health Care*, *25*(S2), 9–20.

Langley, A. (1999). Strategies for theorizing from process data. *Academy of Management Review*, *24*(4), 691–710.

Langley, A., & Tsoukas, H. (2017). The SAGE Handbook of process organization studies. Thousand Oaks, CA: Sage.

Langley, A. N. N., Smallman, C., Tsoukas, H., & Van de Ven, A. H. (2013). Process studies of change in organization and management: Unveiling temporality, activity, and flow. *Academy of Management Journal*, *56*(1), 1–13.

Latour, B. (1987). Science in action: How to follow scientists and engineers through society. Cambridge, Mass., Harvard University Press.

Leviton L. C., Hughes E. F. (2016). Research On the Utilization of Evaluations. *Evaluation Review,* 5(4), 525–548

Leung, L. (2015). Validity, reliability, and generalizability in qualitative research. *Journal of Family Medicine and Primary Care*, *4*(3), 324.

Lilford, R. J., Foster, J., & Pringle, M. (2009). Evaluating eHealth: how to make evaluation more methodologically robust. *PLoS Medicine*, *6*(11), e1000186.

Lincoln, Y. S., & Guba, E. E. (1986). Research, evaluation, and policy analysis: Heuristics for disciplined inquiry. *Review of Policy Research*, *5*(3), 546–565.

Lipsey, M. W. (1993). Theory as method: Small theories of treatments. *New Directions for Evaluation*, *1993*(57), 5–38.

Lomas, J. (2007). The in-between world of knowledge brokering. *BMJ*, *334*(7585), 129–132.

Mark, M. M., Henry, G. T., & Julnes, G. (2000). *Evaluation: An integrated framework for understanding, guiding, and improving policies and programs*. San Francisco: Jossey-Bass.

Mark, M. M. (2003). Toward an integrative view of the theory and practice of program and policy evaluation. *Evaluating social programs and problems: Visions for the new millennium*, 183–204.

Mark, M. M., Greene, J. C., & Shaw, I. (2006). The evaluation of policies, programs and practices. *Handbook of Evaluation*. Thousand Oaks, CA: Sage.

Marshall, G. N., & Hays, R. D. (1994). The patient satisfaction questionnaire short-form (PSQ-18).

Mathews, S. C., McShea, M. J., Hanley, C. L., Ravitz, A., Labrique, A. B., & Cohen, A. B. (2019). Digital health: a path to validation. *NPJ Digital Medicine*, *2*(1), 1–9.

McNair, J. B. (2016). Theoretical basis of health IT evaluation. *Evidence-Based Health Informatics: Promoting Safety and Efficiency Through Scientific Methods and Ethical Policy*, *222*, 39.

Mechael, P., Kaonga, N. N., Chandrasekharan, S., Prakash, M. P., Peter, J., Ganju, A., & Murthy, N. (2019). The elusive path toward measuring health outcomes: lessons learned from a pseudo-randomized controlled trial of a large-scale mobile health initiative. *JMIR mHealth and uHealth*, *7*(8), e14668.

Mohr, D. C., Cheung, K., Schueller, S. M., Brown, C. H., & Duan, N. (2013). Continuous evaluation of evolving behavioral intervention technologies. *American Journal of Preventive Medicine*, *45*(4), 517–523.

Mookherji, S., Mehl, G., Kaonga, N., & Mechael, P. (2015). Unmet need: improving mHealth evaluation rigor to build the evidence base. *Journal of Health Communication*, *20*(10), 1224–1229.

Mueser, K. T., Bond, G. R., Drake, R. E., & Resnick, S. G. (1998). Models of community care for severe mental illness: a review of research on case management. *Schizophrenia Bulletin*, *24*(1), 37–74.

Nicolini, D., Powell, J., Conville, P., & Martinez-Solano, L. (2008). Managing knowledge in the healthcare sector. A review. *International Journal of Management Reviews*, *10*(3), 245–263.

Nilsen, W., Kumar, S., Shar, A., Varoquiers, C., Wiley, T., Riley, W. T., Pavel, M., & Atienza, A. A. (2012). Advancing the science of mHealth. *Journal of Health Communication*, *17*(sup1), 5–10.

Norman, C. D., & Skinner, H. A. (2006). eHEALS: the eHealth literacy scale. *Journal of Medical Internet Research*, *8*(4).

Nutley, S., Walter, I., Davies, H. (2007). *Using evidence: How research can inform public services*. Bristol, England: The Policy Press.

Nykänen, P., Brender, J., Talmon, J., de Keizer, N., Rigby, M., Beuscart-Zephir, M. C., & Ammenwerth, E. (2011). Guideline for good evaluation practice in health informatics (GEP-HI). *International Journal of Medical Informatics*, *80*(12), 815–827.

Orrell, M., & Geraldine, H. (2004). *CANE: Camberwell Assessment of need for the Elderly*. London: Gaskell.

Orlikowski, W. J., & Baroudi, J. J. (1991). Studying information technology in organizations: Research approaches and assumptions. *Information Systems Research*, *2*(1), 1–28.

Pagliari, C., Sloan, D., Gregor, P., Sullivan, F., Detmer, D., Kahan, J. P., Oortwijn, W. & MacGillivray, S. (2005). What is eHealth (4): a scoping exercise to map the field. *Journal of Medical Internet Research*, *7*(1).

Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N., & Hoagwood, K. (2015). Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. *Administration and Policy in Mental Health and Mental Health Services Research*, *42*(5), 533–544.

Parasuraman, A., Zeithaml, V. A., & Berry, L. (1988). SERVQUAL: A multiple-item scale for measuring consumer perceptions of service quality. *Journal of Retailing*, *64*(1), 12–40.

Pham, Q., Wiljer, D., & Cafazzo, J. A. (2016). Beyond the randomized controlled trial: a review of alternatives in mHealth clinical trial methods. *JMIR mHealth and uHealth*, *4*(3), e107.

Porter M. E., & Teisberg E. O (2006). *Redefining competition in health care*. Boston: Harvard Business School Press.

Prasad, P. (1993). Symbolic processes in the implementation of technological change: A symbolic interactionist study of work computerization. *Academy of Management Journal*, *36*(6), 1400–1429.

Proudfoot, J., Klein, B., Barak, A., Carlbring, P., Cuijpers, P., Lange, A., Ritterband, L. & Andersson, G. (2011). Establishing guidelines for executing and reporting internet intervention research. *Cognitive Behaviour Therapy*, *40*(2), 82–97.

Ramirez, R. (1999). Value co-production: intellectual origins and implications for practice and research. *Strategic Management Journal*, *20*(1), 49–65.

Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC

Reynolds, T., Thornicroft, G., Abas, M., Woods, B., Hoe, J., Leese, M., & Orrell, M. (2000). Camberwell Assessment of Need for the Elderly (CANE): Development, validity and reliability. *The British Journal of Psychiatry*, *176*(5), 444–452.

Rigby, M., Ammenwerth, E., Beuscart-Zephir, M. C., Brender, J., Hyppönen, H., Melia, S., Nykänen, P., Talmon, J., & de Keizer, N. (2013). Evidence Based Health Informatics: 10 years of efforts to promote the principle. *Yearbook of Medical Informatics*, *22*(01), 34–46.

Ring, P. S., & Van de Ven, A. H. (1994). Developmental processes of cooperative interorganizational relationships. *Academy of Management Review*, *19*(1), 90–118.

Robertson, A., Cresswell, K., Takian, A., Petrakaki, D., Crowe, S., Cornford, T., Barber, N., Fernando, B., Jacklin, A., & Prescott, R. (2010). Implementation and adoption of nationwide electronic health records in secondary care in England: qualitative analysis of interim results from a prospective national evaluation. *British Medical Journal*, *341*, c4564.

Rogers, E. M. *(*2003*)*. *Diffusion of innovations (*5th ed.*)*. New York: Free Press.

Runz-Jørgensen, S. M., Schiøtz, M. L. & Christensen, U. (2017). Perceived value of eHealth among people living with multimorbidity: a qualitative study. *Journal of Comorbidity*, *7*(1), 96–111.

Røvik, K. A. (2007). *Trender og translasjoner. Ideer som former det 21. århundrets organisasjoner,* Universitetsforlaget, Oslo.

Sager, M., & Zuiderent-Jerak, T. (2021). Knowing Times: Temporalities of Evidence for Implantable Cardioverter Defibrillators. *Science, Technology, & Human Values*, 46(3), 628–654.

Saldaña, J. (2003). *Longitudinal qualitative research: Analyzing change through time*. Walnut Creek, CA: AltaMira Press.

Saldaña, J. (2015). *The coding manual for qualitative researchers*. London, UK: SAGE.

Schwarz, G., & Stensaker, I. (2014). Time to take off the theoretical straightjacket and (re-) introduce phenomenon-driven research. *The Journal of Applied Behavioral Science*, *50*(4), 478–501.

Scriven, M. (1998). Minimalist theory: The least theory that practice requires. *American Journal of Evaluation*, *19*(1), 57–70.

Shadish, W. R. (1998). Evaluation theory is who we are. *American Journal of Evaluation*, *19*(1), 1–19.

Shaw, T., McGregor, D., Brunner, M., Keep, M., Janssen, A., & Barnet, S. (2017). What is eHealth (6)? Development of a Conceptual Model for eHealth: Qualitative Study with Key Informants. *Journal of Medical Internet Research*, *19*(10).

Shulman, K. I. (2000). Clock-drawing: is it the ideal cognitive screening test? *International Journal of Geriatric Psychiatry*, *15*(6), 548–561.

Stufflebeam, D. (2001). Evaluation models. *New directions for evaluation*, *2001*(89), 7–98.

Swinkels, I. C. S., Huygens, M. W. J., Schoenmakers, T. M., Nijeweme-D'Hollosy, W. O., Van Velsen, L., Vermeulen, J., Schoone-Harmsen, M., Van Schayck, O. C., & de Witte, L. (2018). Lessons Learned From a Living Lab on the Broad Adoption of eHealth in Primary Health Care. *Journal of Medical Internet Research*, *20*(3).

Talboom-Kamp, E.P ., Verdijk, N. A., Harmans, L. M., Numans, M. E. and Chavannes, N. H. (2016). An eHealth platform to manage chronic disease in primary care: an innovative approach. *Interactive Journal of Medical Research*, *5*(1), p.e5.

Talmon, J., E. Ammenwerth, J. Brender, N. De Keizer, P. Nykänen and M. Rigby (2009). STARE-HI – Statement on reporting of evaluation studies in Health Informatics. *International Journal of Medical Informatics, 78*(1), 1–9.

Tidd, J., & Bessant, J. (2014). *Strategic innovation management*. John Wiley & Sons.

Tomlinson, M., Rotheram-Borus, M. J., Swartz, L., & Tsai, A. C. (2013). Scaling up mHealth: where is the evidence? *PLoS Med*, *10*(2), e1001382.

Trochim, W. M. (1989). Outcome pattern matching and program theory. *Evaluation and Program Planning*, *12*(4), 355–366.

Urueña, A., Hidalgo, A., & Arenas, Á. E. (2016). Identifying capabilities in innovation projects: Evidences from eHealth. *Journal of Business Research*, *69*(11), 4843–4848.

Van Aartsengel, A., & Kurtoglu, S. (2013). *Handbook on Continuous Improvement Transformation*. Springer Berlin Heidelberg.

Van Gemert-Pijnen, J. E., Nijland, N., van Limburg, M., Ossebaard, H. C., Kelders, S. M., Eysenbach, G., & Seydel, E. R. (2011). A holistic framework to improve the uptake and impact of eHealth technologies. *Journal of Medical Internet Research*, *13*(4).

Van der Vaart, R., van Deursen, A. J., Drossaert, C. H., Taal, E., van Dijk, J. A., & van de Laar, M. A. (2011). Does the eHealth Literacy Scale (eHEALS) measure what it intends to measure? Validation of a Dutch version of the eHEALS in two adult populations. *Journal of Medical Internet Research*, *13*(4).

Van Maanen, J., Sørensen, J. B., & Mitchell, T. R. (2007). The interplay between theory and method. *Academy of Management Review*, *32*(4), 1145–1154.

Vangen, S., & Huxham, C. (2011). The tangled web: unraveling the principle of common goals in collaborations. *Journal of Public Administration Research and Theory*, *22*(4), 731–760.

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, *27*, 425–478.

Zuiderent-Jerak, T., & Bruun Jensen, C. (2007). Editorial introduction: Unpacking 'intervention'in science and technology studies. *Science as Culture*, *16*(3), 227–235.

Weiss, C. H. (1999). The interface between evaluation and public policy. *Evaluation*, *5*(4), 468–486.

White, B. K., Burns, S. K., Giglia, R. C., & Scott, J. A. (2016). Designing evaluation plans for health promotion mHealth interventions: a case study of the Milk Man mobile app. *Health Promotion Journal of Australia: Official Journal of Australian Association of Health Promotion Professionals*, *27*(3), 198–203.

Whittaker, R., Merry, S., Dorey, E., & Maddison, R. (2012). A development and evaluation process for mHealth interventions: examples from New Zealand. *Journal of Health Communication*, *17*(suppl 1), 11–21.

Williams, G., & Doughty, K. (2007). Towards an evaluation framework for telecare services. *Journal of Assistive Technologies*, *1*(1), 42–47.

Wyatt J. (2016). Evidence-based Health Informatics and the Scientific Development of the Field. In: Ammenwerth E, Rigby M, editors. *Evidence-based health informatics: Promoting safety and efficiency through scientific methods and ethical policy*. Amsterdam, Netherlands: IOS Press; 14–24.

Wæraas, A., & Sataøen, H. L. (2014). Trapped in conformity? Translating reputation management into practice. *Scandinavian Journal of Management*, *30*(2), 242–253.

World Health Organization. (2016). Call to Action on Global eHealth Evaluation – Consensus Statement of the WHO Global eHealth Evaluation Meeting, Bellagio,

http://www.ghdonline.org/uploads/The_Bellagio_eHealth_ Evaluation_Call_to_Action-Release.docx.
World Health Organization. (2018). Classification of digital health interventions v1.0: a shared language to describe the uses of digital technology for health.
Yin, R. K. (1989). *Case study research*. Newbury Park, CA: SAGE Publications.
Yin, R. K. (2006). Mixed methods research: Are the methods genuinely integrated or merely parallel. *Research in the Schools*, *13*(1), 41–47.
Yin, R. K. (2015). *Case study research. Design and Methods*. London, UK: SAGE Publications.
Young R. C. (2015). Value-based Cancer care. *The New England Journal of Medicine*, 373, 2593–2595.
Yusof, M. M., Papazafeiropoulou, A., Paul, R. J., & Stergioulas, L. K. (2008). Investigating evaluation frameworks for health information systems. *International Journal of Medical Informatics*, *77*(6), 377–385.

Annex I – Excerpts from data analysis files in Study I.

2016-11-22	Amia sent proposed scales for ICT evaluation to CHI for review	**Defining KPIs**
2016-11-22	Amia sent proposed interview questions for ICT evaluation based on Technology acceptance model to J and M	**Defining qualitative measures for evaluation**
2016-11-23	Sent email to Project managers regarding problem of differences in endpoints and scales between clinical sites.	**Negotiating endpoints and KPIs**
2016-11-23	Contacted ISISEMD project regarding non-valid questionnaires in some countries, non-existing translations, translations by ourselves and how reviewers would react to it.	**Consulting with experts**
2016-11-25	Validating with partners regarding valid translations of different scales in their languages, looking for scales that are valid in all countries	**Negotiating endpoints and KPIs**
2016-11-26	Problem of differences in endpoints and scales and their validations raised on Consortium level	**Negotiating endpoints and KPIs**
2016-11-28	A+J sent a first idea about evaluation of Organization dimension	**Setting strategy of particular domain evaluation**
2016-12-01	Email discussion with clinical partners on which QoL questionnaire to choose	**Negotiating endpoints and KPIs**

Table I. Coding of data

Defining scope	Definition of framework structure. This code applies to various versions of domains: at first, the focus was placed on actor: patient/staff. Later, focus was shifted to domains that were more similar to existing evaluation frameworks in the literature, and also it helped us better divide tasks within the consortium in framework preparation (separated clinical, technological, cost, business, etc.).	***Defining scope and endpoints***	The codes were merged into "Defining scope and endpoints" because "areas of evaluation" and "endpoints" are overlapping to quite big extent. "Areas of evaluation" changed to "scope" as perhaps more scientific word.
Defining endpoints	The code applies to actions that aimed to define clinical endpoints.		
Setting strategy of particular domain evaluation	The code concerns decisions regarding generic principles of the area evaluation. It embodies actions taken to clarify and delimit the scope of the domain since it can be very detailed and specific. BUT this code does not apply to more specific KPI setting actions. For this, another code is applied.	***Defining mechanisms of measurement***	The codes were merged into "Defining mechanisms of measurement" because they all are part of measurement settings: from defining the methods and types of data (quantitative / qualitative), to defining detailed KPIs. Control variables are an important part of a measurement system, as they allow to understand relationships between variables. **Update after group work** on 2017-07-25: "Control variables" in the DECI context mean demographic and background characteristics of the users.
Defining KPIs	The code concerns considerations and definition of actual measurement methods/KPIs in evaluation areas. This code included literature scanning to justify usage of each measurement tool/method.		
Defining qualitative measures for evaluation	The code concerns setting up qualitative measures, such as interview questions for a particular area of evaluation.		
Defining control variables	The code concerns actions that related to control variables such as demographic, socio-economic and other background data of target users.		

Table II. Categorizing (aggregating) of the codes

Category	Q1	Q2	Q3	Q4	Q5	Q6	Q7
Learning approaches from related projects	3						
Acknowledging constraints	1						
Defining evaluation questions					9		
Considering methods of data analysis						1	
Defining measures						27	18
Choosing a methodological approach					5		
Setting-up data collection						22	36
Analyzing feasibility of potential measures				8			
Analyzing stakeholders' perspectives					11		
Setting up monitoring of data collection						6	
Defining expected results							6

Table III. Analysis of the intensity of the categories

DECI process / GEP-HI and AHRQ	*Analyzing*				*Designing*							*Setting-up*		Intermediate*/ Final** version of evaluation framework
	Learning from similar projects	Acknowledging constraints	Analyzing feasibility of potential KPIs	Analyzing stakeholders' perspective	Choosing a methodological approach	Defining evaluation questions	Defining measures	Considering methods of data analysis	Defining expected results	Consulting with experts	Negotiating and localizing	Setting up data collection	Setting up monitoring of data collection	
Result of study design														
Formal acceptance to proceed to the next phase														
Phase 3: Operationalization of methods														
Study type														
Approach														
Assumptions and feasibility assessment														
Frame of reference														
Timing														
Justification of the methodological approach *15. Consider the Impact of Study Design on Relative Cost and Feasibility*														
Expertise														
Outcome measures *5. Consider Both Quantitative and Qualitative Measures* *13. Choose the Measures You Want to Evaluate* *16. Choose Your Final Measures*														
Avoiding Bias														
Quality control on data (measures)														

Table IV. An extract from a comparison between the Health Information Technology Evaluation Toolkit (AHRQ) and Guideline for Good Evaluation practice in Health Informatics (GEP-HI) and the Digital Environment for Cognitive Inclusion (DECI) evaluation process