THESIS FOR THE DEGREE OF LICENTIATE OF ENGINEERING

Planning of a holistic summative eHealth evaluation: The interplay between standards and reality

MONIKA JURKEVICIUTE

Department of Technology Management and Economics

CHALMERS UNIVERSITY OF TECHNOLOGY

Gothenburg, Sweden 2018

Planning of a holistic summative eHealth evaluation: The interplay between standards and reality
MONIKA JURKEVICIUTE

Planning a holistic summative eHealth evaluation: The interplay between standards and reality

MONIKA JURKEVICIUTE

Department of Technology Management and Economics

Chalmers University of Technology

Abstract

Summative evaluation assesses outcomes, produces evidence, and advances knowledge of eHealth implementations in health care provisions. Therefore, its quality is essential to reap benefits from the results generated by evaluation studies. Standardization is considered as among the solutions to improve quality of evaluations as it creates methodological uniformity and enhances research credibility. However, standards are used insufficiently in eHealth evaluation practice.

The purpose of this thesis is to study the use of standards in eHealth evaluation planning practice, in order to provide insight to support knowledge development. This thesis takes a stance in research concerning eHealth evaluation planning with regards to the standardization, translation, and collaboration. This thesis is built on a systematic literature review assessing the extent to which eHealth evaluation frameworks are used in summative eHealth evaluations and a single case study of eHealth evaluation of a project concerning eHealth implementation within several health care contexts. It focuses on the evaluation planning process, assesses adequacy of eHealth evaluation planning guidelines to practice, and identifies reasons that hinder the use of and adherence to standards.

The thesis confirms that eHealth evaluation frameworks are not used in the empirical eHealth evaluations. In contrast, the frameworks and the evaluation planning guidelines are found to be adequate and beneficial to practice. The reasons hindering the use of standards and affecting adherence to them are the insufficient evaluator's experience and resources using a standard, evaluator's unawareness of a standard, inadequacy of a standard to address a target population or a disease, non-existence of a validated version of a standard in a particular location, and a lack of fit between a standard and a scope of the evaluation. A model is developed, suggesting that standards can be viewed as objects translated in a specific context and influenced by collaborative activities. The thesis suggests that adherence to standards in eHealth evaluation practice could be seen as a range that is caused by trade-offs made when standards are translated based on reality.

List of appended papers

Paper 1: Jurkeviciute, M., Lettieri, E., Enam, A., Torres-Bonilla, J., Eriksson, H., Hellström, A. (2018), "Assessing eHealth initiatives: Do theoretical frameworks matter in practice?"

The peer-reviewed paper was presented at the International Conference on Medical Informatics Europe 2018 (MIE2018) in April 2018 in Gothenburg, Sweden.

Contributions: Jurkeviciute was the main author, initiated the study, conducted analysis, and wrote the paper. Lettieri contributed to writing the paper and scientific advisory. Enam and Torres-Bonilla collected data. Eriksson and Hellström contributed to scientific advisory.

Paper 2: Jurkeviciute, M., Enam, A., Torres-Bonilla, J., Eriksson, H. (2018) "Planning a holistic summative eHealth evaluation. Propositions for guideline use and development," submitted to an international journal.

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

Paper 3: Jurkeviciute, M. (2018) "Standards as applied in reality: A case study on the translation of standards in eHealth evaluation practice," submitted to an international journal.

Table of Contents

1 INTRODUCTION

1.1 Background

The rise of the internet has transformed many sectors of economy and industries. Since 2000, health care worldwide has harnessed the potential of the internet and digital technologies (Eysenbach, 2001). Developments in information technology have opened opportunities and scenarios to transform health care and tackle its triple aim: (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 health care is eHealth. It represents various types of information and communication technology (ICT) that are employed in health care provisions, and is considered an effective solution for improving health care services (Khoja et al., 2013; Swinkels et al., 2018). According to the report of the World Health Organization (WHO, 2016), more than 50% of its member states defined national strategy to support eHealth diffusion, and 83% reported to have implemented a number of eHealth initiatives. The diffusion of eHealth in the daily practice of health care systems is slower than expected (Greenhalgh et al., 2017; Swinkels et al., 2018). Decision-makers in health care are particularly careful and need credible evidence when choosing to invest in eHealth, as a wrong choice can lead to harmful events or other negative effects (e.g., Koppel et al., 2005; Han et al., 2006).

The WHO and International Medical Informatics Association (IMIA) emphasized the importance of *summative evaluation* (Lilford et al., 2009) in assessing worth, 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, to support decision-making by people responsible for eHealth implementation and policies, evaluation quality is key to utilize the results produced by evaluation studies (Mookherji et al., 2015). Quality in eHealth evaluation is achieved when a study methodology is based on scientific methods and standardized approaches, a comparison group is used, randomization procedures are in place, sample size calculation applies systematic procedures, data collection is conducted at baseline and end-line, and evaluators are independent (Mookherji et al., 2015).

In 2006, the Evaluation Gap Working Group (Savedoff et al., 2006) cautioned that many evaluations overlook opportunities to collect and analyze various outcomes from eHealth implementations that are essential to fund effective programs, such as health, organizational, and cost outcomes. In a comprehensive review, De Keizer and Ammenwerth (2008) argued that methods used to capture the outcomes of the data for summative evaluations are described insufficiently, implying that it is not clear whether the evaluators used validated measures or developed custom approaches. A recent survey concerning eHealth evaluation studies reported

that the quality of the studies remains insufficient (Mookherji et al., 2015). Two key deficiencies were outlined. First, most evaluations of eHealth assess a narrow range of outcome variables, mostly focusing on behavior change, attitudes, intentions, and cost, and leaving out health outcomes and quality of care. Second, scientific rigor of evaluation is insufficient. Therefore, there is a need to better understand the problems surrounding quality of evaluation studies to increase credibility in the evaluation outcomes.

1.2   Problem, purpose, and research questions

Standardization of summative eHealth evaluations is considered among the possible pathways to increase quality (Ammenwerth, 2004; Greenhalgh et al., 2004; Proudfoot et al., 2011; Janssen et al., 2013a; Mookherji et al., 2015; Cowie et al., 2016). In this thesis, a standard is understood as a non-mandatory directive (i.e., consensus standard) (Allen and Sriram, 2000) and an instrument of control, facilitating coordination and communication (Brunsson and Jacobsson, 2000). Usage of standards, such as guidelines, evaluation frameworks, and standardized metrics, leads to various degrees of methodological uniformity between different studies and enhances generalizability and trust in the research findings (Ammenwerth et al., 2004; Lilford et al., 2009; Mookherji et al., 2015; Cowie et al., 2016).

A need for more standards and training to use them was revealed in a recent survey of eHealth evaluation practitioners (Mookherji et al., 2015). In addition, organizations, such as the European Society for Cardiology, set a goal in their action plan "to develop evaluation standards/criteria for electronic tools; to develop guidelines on the proper conduct of e-health studies and implementation of e-health" (Cowie et al., 2016). For these reasons, numerous scholars have invested time and financial resources to develop standards for a summative eHealth evaluation (e.g., Nykänen et al., 2011; Kidholm et al., 2012; Greenhalgh et al., 2017).

Currently, certain standards for eHealth evaluation are mandatory when publishing results of a study (i.e., Consolidated Standards of Reporting Trials [CONSORT]) (Eysenbach and Consort-EHEALH Group, 2011). The use of other standards (e.g., guidelines for eHealth evaluation planning, evaluation frameworks and standardized metrics) remains voluntary; in practice, the standards are used insufficiently (Greenhalgh et al., 2004; Ammenwerth, 2004; Proudfoot et al., 2011; Janssen et al., 2013a; Mookherji et al., 2015). The insufficient use of standards creates a gap between research and practice in eHealth evaluation, that is, although the use of standards is promoted by different scholars and organizations, standards are seemingly not always used in empirical evaluation studies. Previous research has focused mostly on creation or improvement of different standards (e.g., a newly created evaluation framework by Greenhalgh et al., [2017]). However, the actual use of standards in eHealth evaluation was not addressed in the past years, which is contrary to the emphasis on standards. Previous research has also approached application of standards as their use in original content; thus, overlooking the

possibility to view the use of standards as translation (Røvik, 2007; Czarniawska and Sevón, 2011), which implies adaptation of an object or idea to be suitable to people's interests (Latour, 1987). The use of other standardized improvement concepts in health care 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).

Therefore, the purpose of this thesis is to study the use of standards in eHealth evaluation planning practice, in order to provide insight to support knowledge development.

Following this purpose, this thesis is guided by three research questions.

Research Question 1:

In previous research, scholars have indicated that evaluation practitioners use the standards insufficiently in eHealth evaluation studies (Greenhalgh et al., 2004; Ammenwerth, 2004; Proudfoot et al., 2011; Janssen et al., 2013a; Mookherji et al., 2015). Although standards appear to be used insufficiently, a survey of eHealth evaluation practitioners (Mookherji et al., 2015) concluded that practitioners need them. Following the recommendation by Merton (1987), stating that verifying whether a phenomenon exists is advisable before attempting to explain it, I set out to explore the extent of use of summative eHealth evaluation frameworks (as one type of standards) in the published research articles. To my knowledge, no study has undertaken this endeavor. This will set the ground for the thesis and research in the use of standards in eHealth evaluation practice.

> RQ1: To what extent are standard eHealth evaluation frameworks used in practice?

Research Question 2:

Certain researchers of standardization emphasize that not all standards may be credible or useful (Brunsson and Jacobsson, 2000; Moher et al., 2011). They argue that the development of a standard at times lacks transparency and robustness. Attention should be paid to the basis on which the standard was built and who participated in the development process. Additionally, standards are often developed based on personal preferences (suiting authors' purposes) and authors' experiences, which was highlighted by Moher et al. (2011). They reviewed one type of standards, that is, reporting guidelines for health research, systematically, and concluded that the guidelines were short of evaluation of their effectiveness and their development processes were implicit. Moreover, the importance of evaluating, comparing, and improving existing standards was emphasized by the European Society of Cardiology in their eHealth action plan (Cowie et al., 2016). Implementation of these suggestions has not been conducted by previous research.

Therefore, reflecting upon the adequacy (i.e., to fit the purpose and needs and being of good quality [Hornby et al., 1974]) of a set of standards was deemed necessary in a case study.

> RQ2: To what extent are standards adequate in a practical eHealth evaluation setting, and how can the standards be improved?

Research Question 3:

Standards can help evaluation practitioners enhance quality of evaluation studies, as they create conditions for common denominators between different studies and increase trust and generalizability of research findings (Lilford et al., 2009). Although standardization is increasingly important, standards are used insufficiently in summative eHealth evaluations (Proudfoot et al., 2011; Janssen et al., 2013a; Mookherji et al., 2015; Cowie et al., 2016). Previous research has suggested that evaluations can be affected by culture, assumptions, values, and agendas of evaluators (Stufflebeam, 2001; Chouinard and Cousins, 2009). Other scholars have claimed that evaluation studies can be burdened by 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). Use of standards in evaluation can also be a subject to challenges related to a multi-party research collaboration. Moreover, standards can be seen as translated rather than applied directly in their original content. Hence, understanding what affects the translation of standards in summative eHealth evaluations is beneficial. Accordingly, the third research question is:

> RQ3: What factors can affect the use and translation of standards in a summative eHealth evaluation with multiple collaborating parties?

1.3 Structure of the thesis

This thesis starts with a description of a frame of reference, including concepts related to eHealth evaluation, such as eHealth, evaluation, and standards, and relevant theories used in the studies, particularly the translation theory and model for inter-organizational cooperation. In Chapter 3, the research process, context, and methodology are outlined. Then, research methods used in each study are explained. Research quality issues are discussed as the last element in Chapter 3. Then, a brief summary of every included paper in this thesis is provided in Chapter 4. The thesis proceeds with a discussion of theoretical and practical research implications in Chapter 5, and ends with conclusions and suggestions for future research.

2 FRAME OF REFERENCE

This chapter introduces key relevant theoretical concepts on eHealth evaluation, and a theoretical background with regard to research described in this thesis.

2.1 eHealth

"eHealth" has become a prominent term since 1999, and was marketed as harnessing the opportunities of the internet and e-commerce within health care (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). Thirty-six different definitions of eHealth were identified in the scientific literature from 1997 to 2003 (Pagliari et al., 2005). Among the most accepted (Pagliari et al., 2005) and earliest definitions of eHealth was published by Eysenbach (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 a systematic literature review, Oh et al. (2005) identified the meaning of the term through the *contexts of its utilization*; they found that the "health" part in the term "eHealth" normally reflected a processual perspective of health care and was not referring to health effects. The "technology" part in the term was found to refer to the tools of assistance, enhancement, and enablement to individual's activities, and was not a replacement of them (Oh et al., 2005). In this thesis, I followed the view on eHealth suggested by Kaplan and Shaw (2004), stating that eHealth is a complex innovation implemented in a socio-technical environment (Kaplan and Shaw, 2004), and implementations of eHealth may demand processes to be reorganized, professionals, re-trained, and individual habits, adjusted.

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 (Shaw et al., 2017). The *health in your hands* domain refers to consumers' own observation of their health data (e.g., 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 (e.g., telehealth, fitness coaching and professional support, integrated care, and social media) can be assigned to this domain. Communication can be multichannel: consumer-professional, professional-professional, and consumer-consumer. The *data enabling health* domain refers to the collection, analysis, and use of big amounts of data (e.g., genomics, predictive and precision health care). The analysis can concern different scopes, from an individual level to the entire population (Shaw et al., 2017), thus enhancing research potential (Khoja et al., 2013). All these eHealth developments are the sources of enormous business opportunities that can be delivered through innovative business models (Shaw et al., 2017).

2.2 Evaluation

The evaluation of a program or policy is a progressively important area of science and practice and usually guided by four main goals: (1) to improve programs and organizations; (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 argue 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 to perform an evaluation properly, according to a theory, is almost always impossible, and failed interpretations can lead to counter-productive outcomes (Stufflebeam, 2001). Additionally, the social reality is so complex that the assumption that one can pre-determine aspects to be assessed is naïve (Stufflebeam, 2001). Moreover, aiming for a theory-based evaluation is, at times, an impractical idea; that is, trying to evaluate every item in 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, an intervention, or a 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.

2.2.1 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." However, "properties" in this definition may not be understood as a mere assessment of IT properties. In the 1990s, the evaluation of technologies in health care was understood as a more complex endeavor than merely an attempt to improve technology based on its performance (Burghgraeve, 1995). Since then, scholars have recognized the importance of evaluating the impact of technology to its users, organizations and the like, and aimed to determine the optimal set of concepts to be assessed (Yusof et al., 2008; van Gemert-Pijnen et al., 2011; Andargoli, 2017), such as clinical outcomes, behavior impact, satisfaction, technology acceptance and usability, organizational impact, and cost effectiveness.

The evaluation of eHealth usually undergoes several stages (see Figure 1): preliminary outline, study design, operationalization of methods, project planning, execution of the evaluation study, and completion of the evaluation study (Nykänen et al., 2011). The undertaking starts with an agreement, followed by determining research methodology and study design, and then ends with a study report.
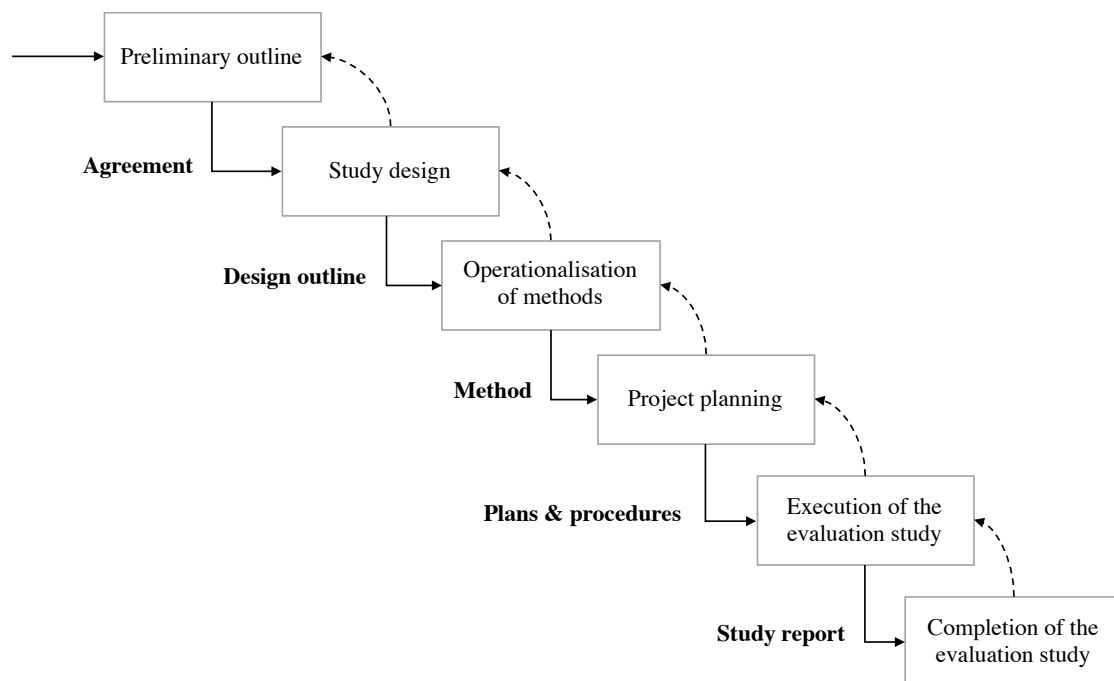

Figure 1. Stages of eHealth evaluation study (Nykänen et al., 2011)

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 (e.g., staff response) during the implementation or design of an eHealth solution. Such evaluation is iterative, producing immediate but less generalizable knowledge. *Summative* evaluation 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 (Lilford et al., 2009). In this thesis, I referred mostly to *summative* evaluation.

Challenges

There are numerous challenges in relation to summative eHealth evaluation's purpose, that is, to generate generalizable knowledge for decision-making regarding eHealth implementation in a health care context (Ammenwerth et al., 2004; Lilford et al., 2009). Several scholars have noted the lack of generalizable knowledge produced by eHealth evaluations (Glasgow, 2007; Proudfoot et al., 2011; Janssen et al., 2013a). Among the challenges that give rise to such lack is insufficiently rigorous research methodologies or study designs (Ammenwerth et al., 2004; Talmon et al., 2009), or incomplete evaluation scope (Glasgow, 2007; Kidholm et al., 2012). Limited by financial resources, evaluators often pursue an insufficient scope of evaluation themes for an informed decision-making, thus reducing the possibility for an eHealth solution to be adopted in real practice (Glasgow, 2007). Another challenge is the lack of a single best method to perform an eHealth evaluation (De Keizer and Ammenwerth, 2008). The research methods and outcome measures are at times tailor-made and context-specific, eliminating the possibility for having common denominators within related studies (Glasgow, 2007). Standardization of the eHealth evaluation research is considered as among the solutions to the lack of generalizable knowledge (Glasgow, 2007; Proudfoot et al., 2011; Janssen et al., 2013a), which will be discussed in more detail in Chapter 2.3.

Another concern in the eHealth evaluation research is the appropriateness of the methodological approach. The choice of methodology depends primarily on the ontological assumption that a researcher follows 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 employing this approach often follow a randomized controlled trial (RCT) design and compose 30% of eHealth evaluation studies (De Keizer and Ammenwerth, 2008). However, scholars have criticized the application of a positivist approach to eHealth evaluation research. They argue that such approach cannot capture the dynamic and socio-technical nature of eHealth (Greenhalgh and Russel, 2010; Robertson et al., 2010). Moreover, they noted the wrong assumption formed in employing this approach, that is, research can be conducted in a controlled environment, supposing that eHealth is embedded in a social context (Greenhalgh and Russel, 2010). 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). 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, as it provides the possibility to ask the "why" questions in addition to the traditional "what," "where," and "who" questions (Lilford et al., 2009). 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 (Glasgow, 2007).

2.3   Standards

Allen and Sriram (2000) identified three categories of standards, namely, de facto, regulatory, and consensus. The category depends on a standard's origin and creation processes. De facto standards are the ones that are widely adopted but not regulated (e.g., a PC keyboard that is defined by the first six characters QWERTY in the upper left side). Regulatory standards are issued by regulatory institutions with a goal of creating uniformity in particular processes of an industry (e.g., standards that regulate safety requirements for particular workplaces or occupations). Consensus standards are issued by local or international bodies to encourage users to conform with a standard but voluntarily (e.g., 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. However, the issuing body of a standard (or a standardizer) takes no formal authority or sanctioning power over the adopters, leaving the adherence to a standard dependent on the free-will of users. What standardizers offer in a standard is only a recipe and guidance to the adopters (Brunsson and Jacobsson, 2000). 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 (1995) 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, in turn, becomes a barrier for accepting a better standard (Farrell and Saloner, 1995; Brunsson and Jacobsson, 2000). Similarly, Farrell and Saloner (1995) argued that standardization can hinder innovation because of users'

potential unwillingness to switch to a new standard.

2.3.1   *Standards in eHealth evaluation*

Because eHealth is a crossroad between the medical, social, and information systems fields, a considerable number of available standardized approaches (i.e., standards) can be applied to eHealth evaluation. To my knowledge, there is currently no scientific publication offering a comprehensive classification of the standards related to eHealth evaluation. The types of standards that this thesis concerns are the following: (1) eHealth evaluation planning guidelines offering guidance in planning an evaluation study; (2) eHealth evaluation frameworks offering a structure in terms of evaluation themes; (3) standard outcome indicators, such as standard scales and questionnaires, designed to measure individual outcomes of a studied intervention; and (4) eHealth evaluation reporting guidelines offering guidance in reporting study results. These types of standards (except when the reporting guidelines refer to the CONSORT standards [Eysenbach and Consort-EHEALH Group, 2011]) are considered *consensus* (voluntary) standards (Allen and Sriram, 2000). Figure 2 depicts the potential use of these types of standards within an eHealth evaluation process.
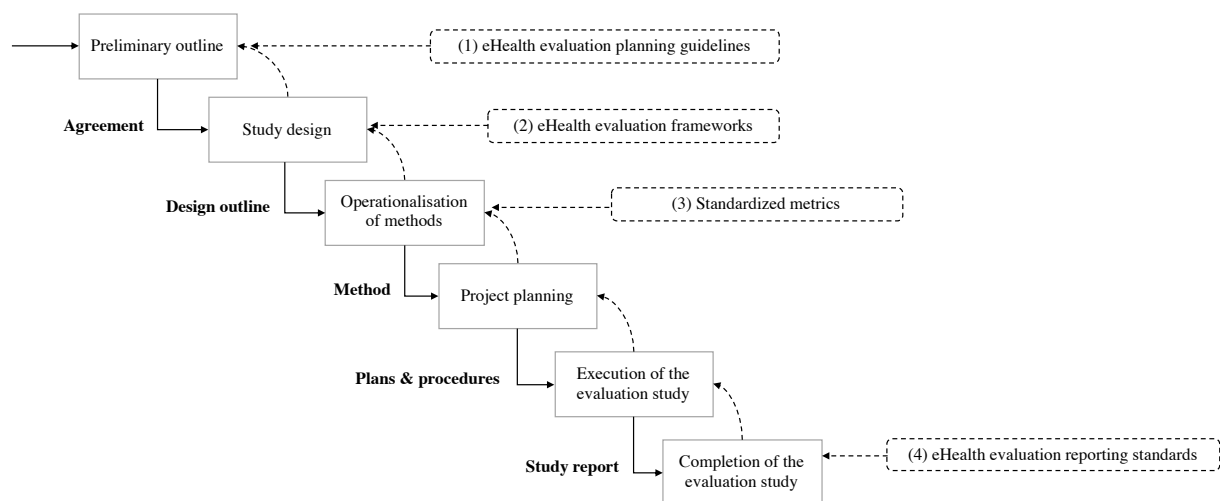
Figure 2. Stages of eHealth evaluation study (Nykänen et al., 2011) supported by standards

(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); and Model for Assessment of Telemedicine Applications, or MAST model (Kidholm et al., 2012). 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, health care 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 (Greenhalgh et al., 2004; Ammenwerth, 2004; Proudfoot et al., 2011; Janssen et al., 2013a), different evaluation frameworks were 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 holistic (mixed-method) 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. Moreover, critics of the standard evaluation frameworks 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 that 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 (e.g., 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 are operationalized through outcome indicators (Nykänen et al., 2011). Numerous standard questionnaires and scales were developed to support measurement of the evaluation themes. Some standards particularly target eHealth (e.g., the eHealth literacy scale [eHEALS; Norman and Skinner, 2006] or eHealth impact questionnaire [Kelly et al., 2015]). Others are generic (e.g., 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. Validity of such standards of measurement of outcome indicators is essential. It increases credibility of the measurement tools that support decision-making regarding the adoption of an eHealth solution. Therefore, different scholars engage in validation of these standards (e.g., validations of eHEALS in Dutch [Van Der Vaart et al., 2011] and Italian [Diviani, 2014] contexts).

(4) Reporting standards on eHealth evaluation

Standards for reporting were created as a result of the observation that the quality of eHealth publications was insufficient (Kaizer and Ammenwerth, 2008). Specifically, scholars did not provide sufficient descriptions of the intervention and methodologies of data collection and analysis. To address such issues, standards for reporting have been suggested. Examples of these are the Consolidated Standards of Reporting Trials (CONSORT statements) (Eysenbach and Consort-EHEALTH Group, 2011) and Statement on Reporting of Evaluation Studies in Health Informatics (STARE-HI) (Talmon et al., 2009). Recently, some journals request to provide a checklist to be filled-in by the authors of a publication, proving that a specific standard has been followed. The CONSORT standards are currently endorsed by over 50% of medical journals. The endorsement of these standards by some of the most prominent journals in the field, such as the Journal of Medical Internet Research, International Journal of Medical Informatics, and American Journal of Preventive Medicine, demonstrate the growing interest in uniformity in reporting.

When different standards are considered to be included 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. The next section 2.4 describes these processes.

2.4 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 process is called translation (Latour, 1987; Røvik, 2007; Czarniawska and Sevón, 2011). Organizational studies have noticed that management ideas change when they are applied 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, namely, copying, addition, omission, and alteration. *Copying* is aimed at accuracy with regard to 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 (e.g., 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). Copying is a likely strategy to be selected when the content is more explicit. In the same way, the abstract content and multiple stakeholders involved make copying less feasible. The selection of a strategy of translation can also depend on the characteristics of a field or industry (Røvik, 2007; Wæraas and Sataøen, 2014).

At times, translation processes do not depend on a single actor, and decision-making on the

outcomes of translation can be a multiparty activity. The following section 2.5 introduces the processes that take place during collaborative activities.

2.5   Interorganizational cooperation

Interorganizational cooperation is a theoretical lens that is relevant in the theoretical framework of this thesis because many evaluations involve multiple actors. With growing international funding in eHealth research and implementation (European Commission's Directorate General for Communications Networks, 2016), studies are becoming more multidisciplinary and international (Greenhalgh and Russel, 2010). Therefore, assessing the impact of interorganizational cooperation in eHealth evaluation research is important. 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 3). These processes involve negotiation, commitment, execution, and assessment. Cooperation happens in iterations, and the outcomes of cooperation are evaluated for reaching its intended goals. The framework extended previous static representations of interorganizational relationships with dynamic elements (Dekker, 2004). To move forward with the cooperation, diverse ideas of different actors need to be aligned (Greenhalgh and Russell, 2010; Vangen and Huxman, 2011). During the *negotiations* stage in the framework (Ring and Van de Ven, 1994), the cooperating parties aim to align and create joint expectations. In this stage, the collaborating parties also evaluate risks of collaboration and establish trust. The goal of this stage is to reach a consensus within the terms of cooperation. When the parties arrive at the *commitments* stage, the consensus regarding the terms of cooperation and the commitment to a future action has been reached. The commitment is manifested in a written document or in a verbal agreement. When the achieved commitments are realized in action, the parties have reached the *executions* stage. At any of the aforementioned phases, a need for potential updates to cooperation can be *assessed* because, among others, changes in the contextual elements or changed positions of the parties can lead to renegotiations (Ring and Van de Ven, 1994).

Negotiations
of joint expectations
risk & trust through
formal bargaining
informal sense-making

Commitments
for future action through

formal legal contract
psychological contract

Assessments
based on:
efficiency
equity

Executions
of commitments through

role interactions
personal interactions
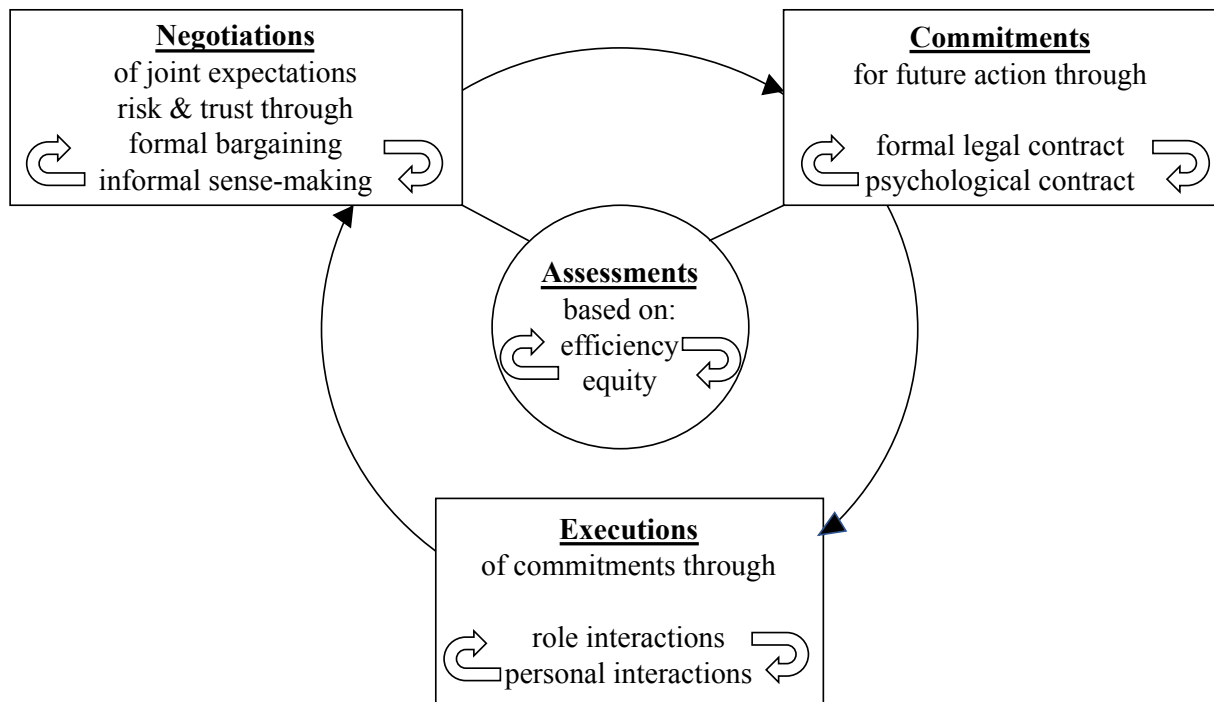
Figure 3. Process framework of the development of cooperative interorganizational relationships (Ring and Van de Ven, 1994)

The framework 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 health care, misses a process view. Such perspective can enrich the understanding of the different elements surrounding the phenomenon, and help improve those elements. In this thesis, the framework is employed in examining the translation of eHealth evaluation standards as enacted through the processes of interorganizational cooperation.

2.6   Summary

An integration of the theories and concepts previously described allows to explore the lack of usage of standards in eHealth evaluation practice. The theories can be summarized in a conceptual theoretical model (see Figure 4). Depending on the evaluation assignment, different standards for eHealth evaluation may be considered during evaluation planning to support the methodology of the study. Consideration of the standards undergoes a translation process that determines the translation strategy employed (Røvik, 2007). The translation process happens in a certain context and can be influenced by different contextual factors (e.g. culture, assumptions, values and agendas of evaluators [Stufflebeam, 2001; Chouinard and Cousins, 2009]). In the case analyzed in this thesis, collaboration between different actors was central. Therefore, the collaboration is depicted as a separate element in the conceptual theoretical model (Figure 4). Challenges emerging from collaborative activities in a multiparty research have been noted in the previous research, e.g. alignment of goals and ideas (Greenhalgh and

Russell, 2010; Vangen and Huxman, 2011). The selected translation strategy determines how a standard will be used, if at all, and whether it will be adhered to in a particular study of summative eHealth evaluation.
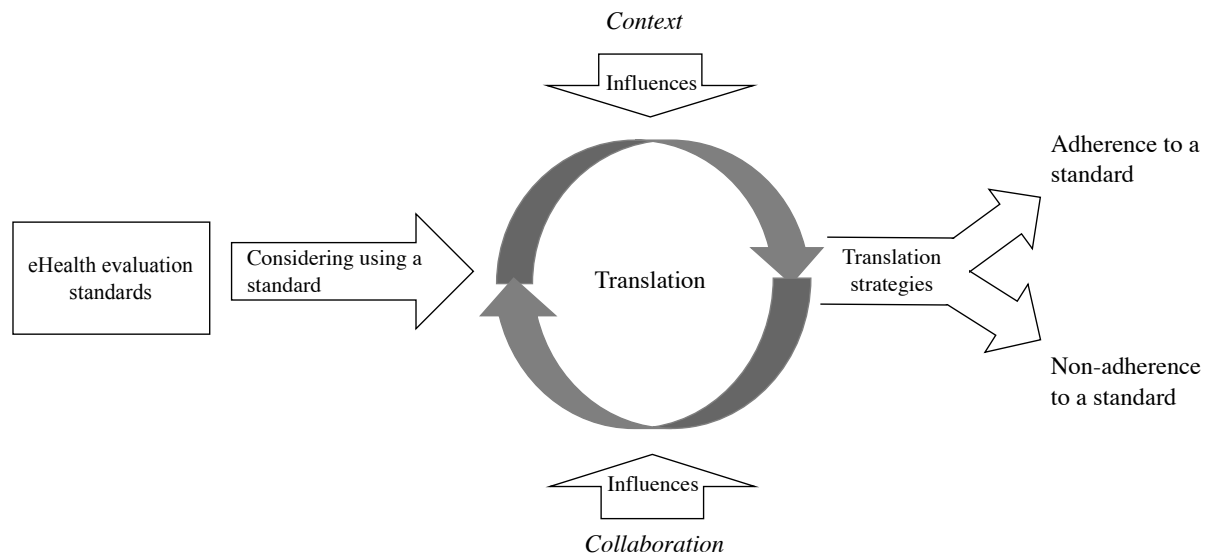
Figure 4. Conceptual model for analysis and discussion

3 METHODOLOGY

3.1 Research process

The research described in this thesis comprised three studies, resulting in three papers. The first study was a systematic literature review performed from 2016 to 2017; it was described in Paper 1. The second study was based on the data collected from the evaluation planning phase in the European Union's (EU's) Digital Environment for Cognitive Inclusion (DECI) project (2015–2017); it was covered in Papers 2 and 3. All three studies were conducted after immersing in an empirical setting (see section 3.2). Additionally, this section describes the research process (see Figure 5), timeline (Figure 6), and a summary of the key elements of the conducted research (Table 1).

```
┌─────────────────────────────────┐
┆ eHealth evaluation planning in the DECI ┆
┆           project                ┆
└─────────────────────────────────┘
                 │
                 ▼
┌─────────────────────────────────┐
│    Observation from practice     │
│ Creating methodological uniformity with │
│   similar studies is problematic │
└─────────────────────────────────┘
                 │
                 ▼
┌─────────────────────────────────┐
┆ Acquaintance with the literature of eHealth ┆
┆           evaluation             ┆
└─────────────────────────────────┘
                 │
                 ▼
┌─────────────────────────────────┐
│    Theoretical problematization  │
│ There is an insufficient methodological │
│  uniformity among eHealth evaluation │
│           studies                │
└─────────────────────────────────┘
                 │
                 ▼
┌─────────────────────────────────┐
│ **Standardization** can enhance quality of evaluation studies and │
│ can create a higher degree of methodological uniformity among │
│ eHealth evaluation studies (Ammenwerth et al., 2004; Lilford et │
│ al., 2009; Mookherji et al., 2015; Cowie et al., 2016) │
└─────────────────────────────────┘
                 │
                 ▼
**RQ1**: To what extent are eHealth evaluation
frameworks used in practice?
                 │
                 ▼
┌─────────────────────────────────┐
│      **Study 1 (paper 1)**       │
│ Assessing the actual use of eHealth │
│ evaluation frameworks in practice │
└─────────────────────────────────┘
```

RQ2: To what extent are standards adequate in a practical eHealth evaluation setting, and how can the standards be improved?

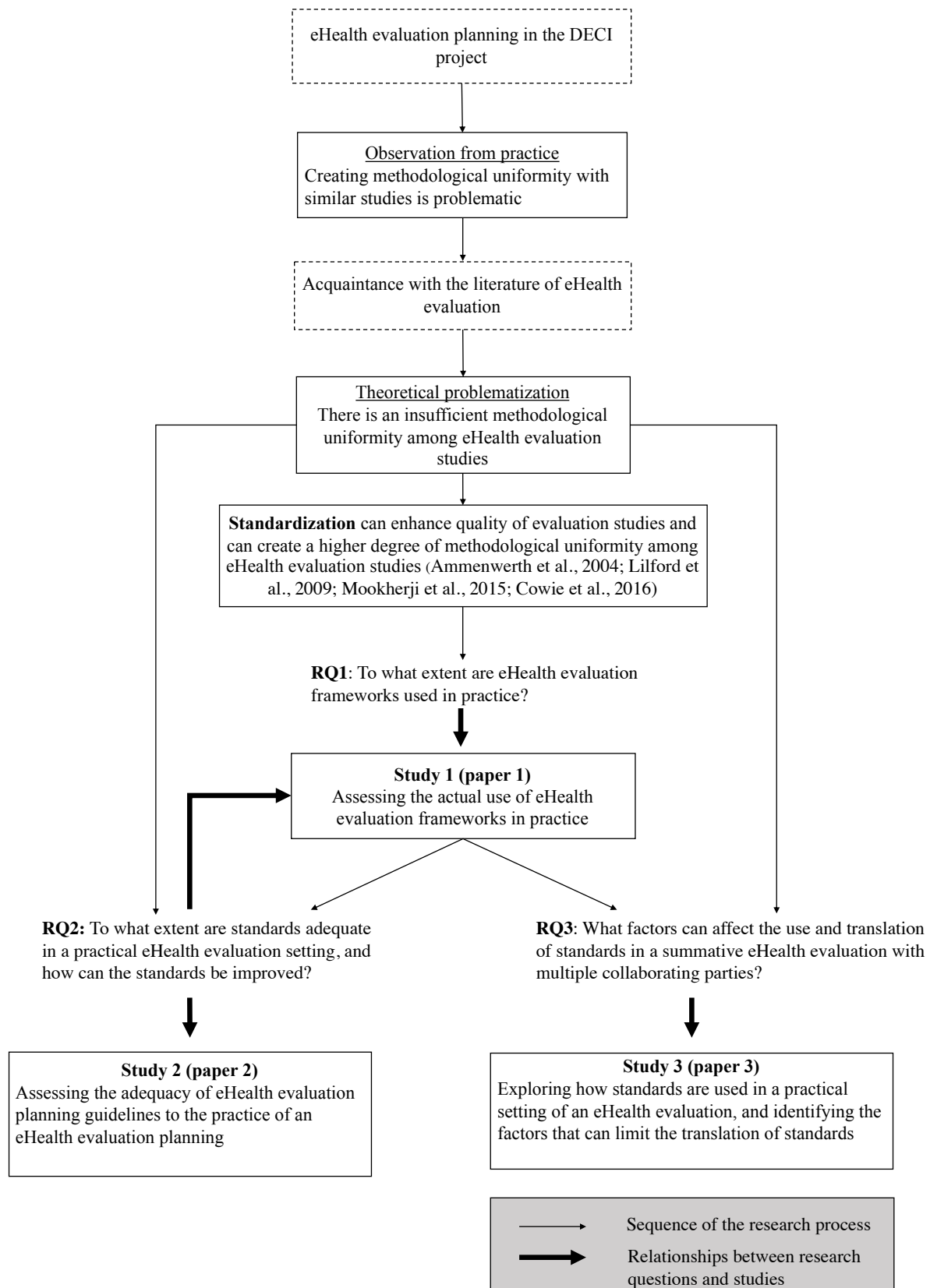**RQ3**: What factors can affect the use and translation of standards in a summative eHealth evaluation with multiple collaborating parties?

```
┌─────────────────────────────────┐      ┌─────────────────────────────────┐
│      **Study 2 (paper 2)**       │      │      **Study 3 (paper 3)**       │
│ Assessing the adequacy of eHealth evaluation │      │ Exploring how standards are used in a practical │
│ planning guidelines to the practice of an │      │ setting of an eHealth evaluation, and identifying the │
│ eHealth evaluation planning      │      │ factors that can limit the translation of standards │
└─────────────────────────────────┘      └─────────────────────────────────┘
```

```
┌──────────────────────────────────────────────┐
│  ──────────►   Sequence of the research process │
│                                                │
│  ━━━━━━►       Relationships between research   │
│                questions and studies           │
└──────────────────────────────────────────────┘
```

Figure 5. Research process

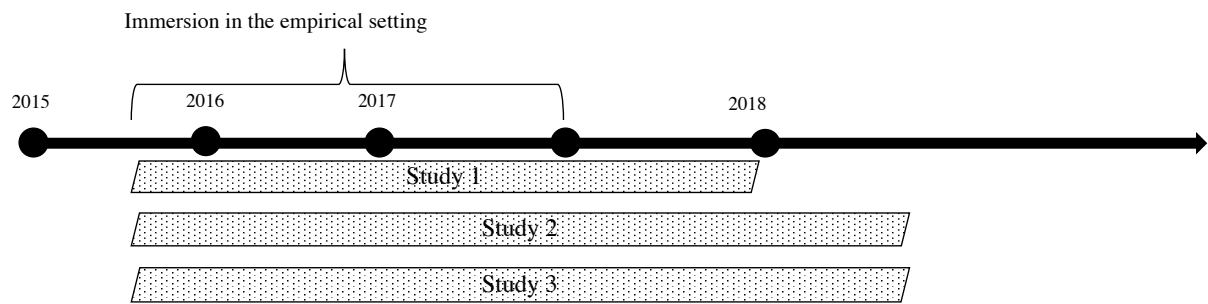Figure 6 presents a timeline of data collection and the three studies.

Figure 6. Research timeline

3.2 Empirical setting

The empirical context was the EU's DECI project, which aimed to define a business model to provide IT-based services to elderly people with mild cognitive impairment (MCI) or mild dementia. The DECI is an eHealth initiative that introduced digital technologies in four different countries and health care contexts. The components of eHealth in the DECI project were: (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 is an indoor sensor to monitor patients' activity; (3) a user coaching and training system, which provides a customizable physical training program to patients; and (4) a cognitive exercise system, which provides a customizable online cognitive stimulation program to patients. Aside from the technology-enabled care, the role of a service coordinator or a case manager is introduced in the health care model. In general, a role of a case manager involves coordination and integration of different cervices into a cohesive program, which is customized individually to meet the needs of patients (Mueser et al., 1998). This new role in the DECI project implies the coordination of the integrated care service to patients and caregivers, provided in each pilot site. By providing the technology-enabled services at the patients' homes, the project also aimed to improve quality of life and increase independence of patients.

The DECI project was selected as the empirical setting for studies 2 and 3 because of convenience in terms of access to data and peculiarities of the project. The DECI was a multinational and multidisciplinary project involving eight partners from Italy, Sweden, Spain, Israel, and the Netherlands (see Figure 7). The multidisciplinary nature of the DECI project entailed continuous interaction between partners of the research consortia. Because partners represented different business types, such as health care organizations, information technology firms, and science institutions, the different perspectives added complexity to cooperation.

Figure 7. Composition of the Digital Environment for Cognitive Inclusion (DECI) research consortium

To enable cross-country comparisons of study outcomes and create premises for transferability of the DECI outcomes to other contexts, the project included a work package related to evaluation of the project outcomes. In this thesis, studies 2 and 3 were built on the evaluation planning processes within the DECI project.

3.3 Reflections on my role in the project and the research

Research in this thesis was based on data collected from the DECI (for studies 2 and 3). Since I took part in the project, it is necessary to reflect upon the potential implications of my involvement in the research. In the DECI project, my role as a researcher was to participate in different work packages, from patient needs assessment to assessing holistically the outcomes of the project. My key responsibility in the project was to coordinate the assessment of the project outcomes. Additionally, I was a researcher in different work packages, such as patient needs assessment and business modelling. I had no previous experience in eHealth, although I was involved in several EU projects.

Several considerations borrowed from the ethnographic research described by Cunliffe and Karunanayake (2013) relate to multiple interconnections that can emerge between the

researcher and the studied subject:

(1) *Insiderness–Outsiderness*, with the most applicable aspect, "Does the researcher have an ongoing role in the research site or work primarily outside the site?"
(2) *Sameness–Difference*, related to the similarity between the researcher and respondents.
(3) *Engagement–Distance*, referring to the researcher's level of engagement in the researched activities, and what part do the respondents take in creating knowledge?
(4) *Political Activism–Active Neutrality*, with the most applicable aspect, "Does the researcher intervene and/or play an active role in the struggles of respondents?

Regarding the *Insiderness–Outsiderness* aspect, I had an active role in the DECI project and was in charge of the work package concerning evaluation activities. Therefore, I was an insider. However, at the time of evaluation planning in the DECI project, the ideas of the research studies discussed in this thesis had not yet emerged. My practical experience in the project revealed the tension between theory and practice in eHealth evaluation and raised curiosity that led me to the research questions.

Regarding the *Sameness–Difference* aspect, I would reflect upon all the participants of the project being rather different owing to different backgrounds and businesses they represent. However, all participants of the consortium tried to develop a common vocabulary and achieve common goals. As there are no respondents in my research, and the data used reflect the organic development of activities during evaluation planning, the sameness–difference aspect is less relevant.

Regarding the *Engagement–Distance* aspect, the activities driven by different partners in the DECI project were not related to the research purposes of this thesis. Hence, at the time of the evaluation planning activities in the DECI project, the partners were not engaged in the research and scientific knowledge generation related to this thesis. Nevertheless, I contributed greatly to the creation of data used in this thesis. However, as discussed previously, the activities that later became the data used in this thesis were conducted for the purposes of the DECI project and not the research of this thesis.

Regarding the *Political Activism–Active Neutrality* aspect, I consider myself as taking an active role in the activities (and "struggles") of the project.

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

3.4   Research design

The purpose and research questions discussed in this thesis concerned the investigation of the use of standards in eHealth evaluation practice. The exploration of the extent of the actual use

of standards in eHealth evaluation practice calls for an analysis of historical data, which, in this case, were published studies of the empirical eHealth evaluations. Hart (1998) suggested that if one needs to "identify relationships between ideas and practice" or "rationalize the significance of the problem," a literature review is an appropriate research design. Hence, a literature review was deemed as a suitable research design for study 1, as it helps evaluate evidence from available documents (published empirical studies) in relation to the research aim (Hart, 1998).

The research conducted in studies 2 and 3 was built on a qualitative single-case study design. Such design allows to study a phenomenon in a unique setting, and dive deeper into the problems (Flick, 2014). A case study design also provides an opportunity for a multifaceted assessment of a setting (Yin, 2014). However, 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 is particularly beneficial, as suggested by Yin (2014). Being actively involved in the DECI project, I acknowledge this "disadvantage" in Chapter 3.3 where I reflected upon my role in the project and research.

Moreover, the DECI project in studies 2 and 3 was explored by employing a process research perspective (Langley, 1999; Langley and Tsoukas, 2017). In an extensive systematic literature review on diffusion of innovations in health care, Greenhalgh et al. (2004) noted that an understanding of the processes surrounding innovations in health care is not only crucial but also highly missing in the published literature. It can help understand how innovations are implemented and adopted by users of an organization. Evaluation is an important part of eHealth implementation projects, with implications concerning evidence for decision-making, which, in turn, can lead to the adoption and sustainability of an eHealth solution. An eHealth evaluation is an activity consisting 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 different approaches to study a process exist (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).

3.5 Research methods

This chapter describes the method of choice, data collection, sampling, and data analysis for each of the three papers that were included in this thesis.

3.5.1 Summary of the conducted research

Table 1 provides a brief overview of the research described in this thesis.

Table 1: Summary of the conducted research

Study	Study 1	Study 2	Study 3
Purpose	To investigate the actual application of the theoretical evaluation frameworks when assessing eHealth initiatives.	To assess the adequacy of eHealth evaluation planning guidelines through the lens of a practical case.	To explore how standards are used in a practical setting of an eHealth evaluation, and to identify the factors that can hinder their use.
Empirical setting	N/A	DECI project	DECI project
Data sources	Published papers on empirical summative eHealth evaluation	E-mail correspondence, minutes of the meeting, electronic versions of the DECI project's evaluation plan	Participants' notes, e-mail correspondence, and minutes of the meeting
Output	Empirical summative eHealth evaluation studies do not rely on theoretical evaluation frameworks and prefer tailor-made approaches, thus reducing the comparability of assessment results.	Confirmed adequacy of the selected guidelines to the eHealth evaluation planning practice. Identified areas for improvement in the guidelines that can enhance comparability of the results generated by different studies. Demonstrated value of adherence to the guidelines in practice.	Usage of a standard could be understood as a range, with different ranges of adherence. Variation in adherence to the standards can create heterogeneity among different studies. Practical circumstances create factors that can hinder the use of standards.

3.5.2 Study 1: Assessing eHealth initiatives: Do theoretical evaluation frameworks matter in practice?

To investigate the actual application of the theoretical evaluation frameworks when assessing eHealth initiatives, a systematic literature review was performed (Tranfield et al., 2003; Pitaway et al., 2004). Figure 8 depicts the process of the systematic literature selection. The period of the search was from 1990 to 2016, which aligned with the advent of the term "eHealth" (Oh et al., 2005). The search strings were "eHealth interventions" AND "research methods"; "eHealth interventions" AND "study design"; and "eHealth interventions" AND "evaluation methods." Among the existing multiple variants of the terms denoting the use of ICT in health care (e.g., "health IT," "health information system," etc.), "eHealth" was the only variant used for the

search to limit the sample. The initial search yielded 1,313 (see Figure 8) hits in the Google Scholar, and 227 in the Scopus databases.

The first screening was conducted by two authors, who scanned individually the types of articles found during the search. As we were in search of the empirical evaluations of eHealth initiatives, we excluded meta-analysis, books, literature reviews, patent descriptions, and pharmacological interventions. The first screening reduced the number of articles to 780, from which duplicates between the databases were eliminated, resulting in 697 that went under further review.

The second stage of screening was based on the review of the article titles. To delimit the sample, we included in the analysis only those articles that seemingly reported empirical results from the evaluations of eHealth interventions. This step yielded 261 articles.

The third stage of screening was based on the abstracts of the articles. The exclusion and inclusion criteria were applied again, as the abstracts provided more details on the articles. Articles must also include empirical data. The remaining sample after this step comprised 90 articles.

Lastly, the first author reviewed the methodologies in the remaining articles to extract those that involved *summative* empirical evaluations (Lincoln and Guba, 1986) of eHealth initiatives, which were the interest of this study. Formative evaluations (Lincoln and Guba, 1986) of eHealth were excluded from the final sample. The selection process produced 23 articles that contained results from empirical summative eHealth evaluations.
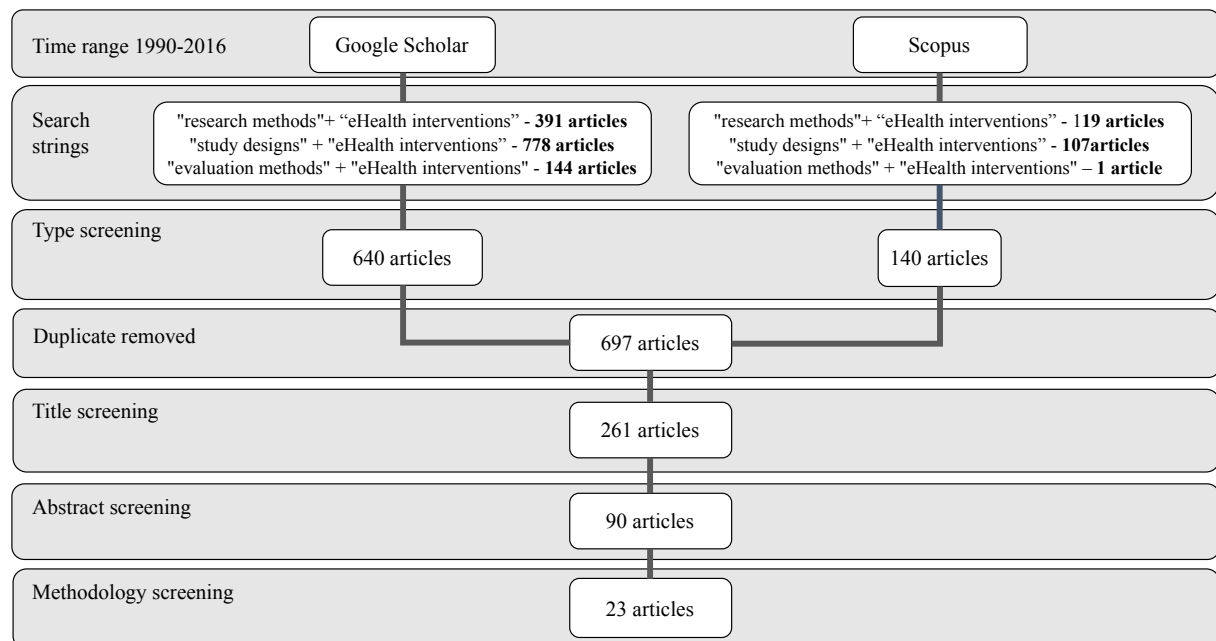

Figure 8. Systematic literature review process

At the first stage of the analysis, we looked for evidence in the selected empirical articles,

checking whether the evaluation questions in the empirical study were supported by *any* published conceptual eHealth evaluation framework. Additionally, a list of references was screened in each article in search for a reference to any conceptual eHealth evaluation framework.

The purpose of the second stage of analysis was to verify whether the empirical studies could have been based on at least one of several of conceptual summative eHealth evaluation frameworks. That is, whether the reported evaluation themes in the empirical studies were among the recommended evaluation themes by a conceptual framework. The conceptual summative evaluation frameworks were selected from two recent literature reviews on eHealth evaluation frameworks (Van Gemert-Pijnen et al., 2011; Andargoli et al., 2017). The selected conceptual evaluation frameworks were the Health Information Technology Reference-based Evaluation Framework, HITREF (Sockolow et al., 2012); evaluation criteria for eHealth services (Hamid and Sarmad, 2008); framework for health technology decisions (Kazanjian and Green, 2002); and Telehealth Evaluation Framework (Hebert, 2001). Two extra conceptual summative evaluation frameworks that did not appear in these literature reviews were added: the Health Technology Assessment Core model (Lampe et al., 2009) and a Model for Assessment of Telemedicine Applications, MAST (Kidholm et al., 2012). These frameworks offered a comprehensive and advanced structure of evaluation themes to consider when planning an eHealth evaluation (Nykänen et al., 2011; Kidholm et al., 2012). During analysis, the reported evaluation themes in the empirical studies were compared with the recommended ones by the conceptual frameworks. A *good match* between the empirical and conceptual frameworks was assumed if the conceptual framework includes all the evaluation themes that have been targeted by the empirical paper; a *partial match*, if the conceptual framework includes at least one evaluation theme (but not all) that were targeted by the empirical paper; a *no match*, if the conceptual framework does not include any of the evaluation themes that were targeted by the empirical paper; and a *not applicable match*, if the conceptual framework is developed at a later date than that of the empirical paper.

3.5.3 *Study 2: Planning a holistic summative eHealth evaluation: Propositions for guideline use and development*

The aim of the study was to assess the adequacy of eHealth evaluation planning guidelines through the lens of a practical case. The "practice" here was reflected through developing an eHealth evaluation planning process in the DECI project and comparing it to the conceptual eHealth evaluation planning guidelines. The guideline selection was based on the criteria that a recommended process or step-by-step guidance is provided in the guideline. The selected guidelines were Cusack et al.'s (2009) Health Information Technology Evaluation Toolkit

(AHRQ) and Nykänen et al.'s (2011) Guideline for Good Evaluation practice in Health Informatics (GEP-HI).

Creating the process map of the DECI evaluation planning

To compare with the guidelines, a process map of evaluation planning from the empirical DECI project had to be produced. Data were collected from the internal documents of the project. The internal documents included e-mail correspondences between project partners (n = 261) (Gehman et al., 2013), minutes of the meeting (n = 8) dated between September 2015 and June 2017 (a consent to use these internal documents was received), and different working versions of the evaluation plan of the DECI project (n = 32). One researcher reviewed and organized all the documents into a chronological timeline, then extracted a total of 301 activities that each document reflected and discussed. After which, data analysis on the empirical process of eHealth evaluation planning in the DECI project began.

In a chronological timeline, a code was assigned to each activity. Each code was defined subjectively by one researcher and reflected the purpose of the activity (Grbich, 2012; Saldaña, 2015). Figure 9 presents the example of the chronological list of activities in the DECI project and their assigned codes.

2016-11-16	Skype call with RRD on technological evaluation: integrated/separate technologies to be evaluated, scales	**Setting strategy of particular domain evaluation**
2016-11-16	Asked FPM for a brief guideline contribution on BM evaluation. Not asking for full-responsibility.	**Noise, no impact**
2016-11-17	Asked RRD for contribution in Cost evaluation	**Noise, no impact**
2016-11-18	Sent CMO examples to S,P,H,A,J	**Considering general methodology of evaluation**
2016-11-21	Asked clinical partners if they need any input to ethical applications from WP6 side	**Noise, no impact**
2016-11-21	Discovered HTA handbook and shared with A,J, H	**Considering general methodology of evaluation**
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**

Figure 9. Coding of data

As the coding was performed by one of the researchers, all the choices of the codes and their definitions were described thoroughly in an analytic memo (Saldaña, 2015). Two other collaborating authors of the study examined the material based on the analytic memo. Disagreements were resolved in a meeting that was documented in the minutes.

However, multiple activities occurred repeatedly during the different periods of data collection (2015–2017) while receiving the same code. Therefore, many codes had been repeated multiple times and thus need to be organized to develop a meaningful process. As such, the codes were aggregated into larger categories (n = 11) based on the summative features of the purpose or activities. When several codes shared a similar purpose (e.g., defining quantitative and qualitative measures), they were aggregated into a category (e.g., "defining mechanisms of

measurement"). Figure 10 presents an extract from the analytic memo depicting the categorization.

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.		

Figure 10. Categorization of codes

To make the guidelines and the DECI process comparable (both conceptual guidelines provided a step-by-step guidance to a different extent), it was deemed appropriate to identify a sequence in the DECI evaluation planning process as well (Langley, 1999). However, after aggregating the codes into categories, it became unclear which sequential place each category could occupy in the process map. When recoding the chronological list of activities using the aggregated categories, it was observed that the categories still repeated in different periods; it would not be reasonable to present this cyclical repetition in a process map. To make each category appear in the process map only once, the data collection period between September 2015 and June 2017 was divided into three months, composing seven periods in total. Then, a number of each category's appearance in a particular period was counted (see Figure 11).

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

Figure 11. Analysis of the intensity of the categories

For each category, the period containing the highest count was identified. It was assumed that the activities represented by the category were critical to a particular quarter, if a category appeared in it in the highest count. Finally, this determined a sequential place of a category in the process map.

Comparison between the guidelines and the process map of the DECI project

The comparison between the guidelines, AHRQ and GEP-HI, and the DECI project was performed independently by three authors using a pattern-matching technique (Trochim, 1989). This technique helps associate theoretical and observed patterns. Differences in outcomes of the comparison were discussed and resolved between the authors. The purposes and activities of the recommended steps in the AHRQ and GEP-HI guidelines were compared with the steps in the DECI evaluation planning process. The AHRQ's full scope and GEP-HI's first four (i.e., preliminary outline, study design, operationalization of methods, and project planning) were considered for this analysis. A "full match" between the theoretical and empirical patterns was assumed when the steps had similar purposes and contained similar activities. A "partial match" was assumed between the steps when the purposes or activities of a step in the guidelines were sub-purposes or sub-activities of a step in the DECI, and vice versa. A "no match" was assumed when both purposes and activities of a particular step in the guidelines did not have an equivalent in the purposes or activities in the empirical process. The comparison between the AHRQ and GEP-HI guidelines and the DECI empirical process is presented in Figure 12. A black color in the table cells represents a "full match;" a grey color, a "partial match;" and a white, a "no match."

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)							■					▨	▨	

Figure 12. 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

3.5.4 Study 3: Standards as applied in reality: A case study on the translation of standards in eHealth evaluation practice

The aim of this study was to explore how standards are used in a practical setting of an eHealth evaluation, and to identify the factors that can hinder their use. This study focused on the evaluation planning activities of the DECI project.

Consistent with the qualitative research and a case study design (Yin, 2014), the data collection aimed at reflecting on the longitudinal events in the project. The data used in this study were derived from minutes of the meeting (n = 8), e-mail conversations (n = 261), and participant observations (notes) obtained between September 2015 and June 2017. At first, the standards for analysis were selected. Data reflecting events and discussions on usage of each standard were outlined in a chronological order using time stamps in the data. This helped identify the evolution of events for each standard, from an idea to use the standard to the decision on its actual use. With an aim to reflect upon various barriers and uses of standards in the DECI project, the standards for this study were selected based on the ability of the data related to certain standards to capture the translation strategies. Hence, four standards were selected for analysis.

For each standard, data were analyzed using the process framework of the development of cooperative interorganizational relationships (Ring and Van de Ven, 1994), and then organized according to the key processes of the framework (i.e., negotiation and renegotiation, commitment, and execution) Additional elements of cooperation, such as uncertainties, expectations, and bargaining (Ring and Van de Ven, 1994), were searched in the data. Uncertainties and bargaining to resolve them were interpreted as barriers to the use of standards in an eHealth evaluation. At the commitment stage of the cooperation processes, a decision on how to apply a standard was taken.

Decisions on the use of standards were interpreted as translation strategies defined by Røvik (2007). A standard was anticipated to be translated using *copying* strategy, if a standard was used with no modifications to its content; *addition* strategy, if the content of a standard was supplemented by extra elements; *omission* strategy, if one or more components were removed from the content of a standard, for certain reasons; and *alteration* strategy, if the standard was altered to a large extent but not in the same way as omission and addition strategies. An example of a change in a methodological logic is when elements of a quantitative standard would be used in a qualitative interview.

3.6   Research quality

Guba and Lincoln (1989) provided a set of quality criteria for research consisting of credibility, transferability, dependability, and confirmability.

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 has to be aware of his/her own constructions (Halldórsson and Aastrup, 2003).

In study 1 (i.e., a literature review), we selected scientific papers of published empirical summative eHealth evaluations through a systematic literature review to ensure objectivity in the paper selection. Papers that met the inclusion and exclusion criteria and were to be included in the analysis were selected individually and discussed by three researchers. After that, the final list of papers was created.

In study 2, the "own constructions" in the development of the DECI evaluation planning process and its comparison with the eHealth evaluation guidelines were validated by three researchers by performing the same analysis and then comparing the results, and discussing and resolving discrepancies.

In study 3, the usage 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."

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 in studies 2 and 3 depended on the empirical context that was related to a holistic 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 studies 2 and 3 can provide reference for analytical generalization (Yin, 1989).

Dependability

Dependability relates to the "audit trail" of the conducted research (Guba and Lincoln, 1994). Dependability emphasizes an importance of the possibility to track the research process, method, and decisions made. For study 1, the analytical procedures of a systematic literature review were described step-by-step in the paper. Three of the authors were involved in the selection of the papers, and the references of the selected papers were defined in the paper. In study 2, the data, analytical reasoning, and decisions taken during the research process were documented in detail in an analytic memo and 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 3 was based on data from e-mail communications between the DECI project partners. 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.

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 taken in studies 2 and 3 to use data from e-mail communications can have limitations. For example, I might not possess all e-mails that had circulated among project partners in relation to the DECI evaluation planning. However, being in charge of the evaluation activities, this would be less likely to occur. In study 3, 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.

4   SUMMARY OF APPENDED PAPERS

This chapter provides a brief introduction to the results of each three papers appended in the thesis. The focus is placed to the outcomes and contributions of each study in relation to the purpose and research questions of the thesis.

4.1   Paper 1: Assessing eHealth initiatives: Do theoretical frameworks matter in practice?

Paper 1 aimed to investigate the actual application of the theoretical evaluation frameworks when assessing eHealth initiatives. Through a systematic literature review, we selected 23 summative eHealth evaluation papers published in the peer-reviewed journals and conference proceedings. References to the standardized eHealth evaluation frameworks (hereinafter referred to as standardized framework) were searched in the publications of summative eHealth evaluations. The results indicated that most reviewed empirical studies did not support their studies with any standardized framework. Only two papers referred briefly to the standardized frameworks, but the evaluation themes differed from the ones advised by the frameworks. Then, we investigated whether the selected standardized frameworks were appropriate for the selected empirical studies, based on evaluation themes in both empirical studies and the standardized frameworks. Evaluation themes reported in the empirical papers were compared with the themes outlined in six published standardized frameworks for summative eHealth evaluations. The analysis showed that 74% of the reviewed empirical papers could have built the assessment on at least one of the analyzed standardized frameworks. Paper 1 indicated that further research is needed to understand the reasons why studies did not apply or refer to standardized frameworks in empirical research. This study empirically validated the stated problem that empirical studies on eHealth evaluations did not support their evaluation with the standardized approaches.

4.2   Paper 2: Planning a holistic summative eHealth evaluation. Propositions for guideline use and development

The aim of the study was to assess the adequacy of eHealth evaluation planning guidelines through the lens of a practical case. An analysis on the similarities and differences between the eHealth evaluation planning guidelines and the empirical DECI case of eHealth evaluation planning was performed in this study. At first, the 11-step process map was developed and then compared with the theoretical guidelines of eHealth evaluation planning. The process map was developed from the retrospective data offered by the Digital Environment for Cognitive Inclusion (DECI) empirical setting. The process map comprised three phases, namely, analyzing, designing, and setting up. Each phase contained different process steps that led to the creation of the evaluation plan for the DECI project.

The findings derived from the comparison between the evaluation planning activities suggested in the AHRQ and GEP-HI guidelines and the DECI empirical process map indicated that the former were quite adequate for eHealth evaluation planning practice but can be improved further. The comparison revealed absent elements in the guidelines and unaddressed activities in the DECI process map. It also indicated that the following two important steps were unaddressed in practice (in the DECI evaluation planning): risk analysis (from the GEP-HI), continuous observation of potential barriers to evaluation, and documentation of lessons learned (from the AHRQ). The guidelines suggested that risks, such as changes in key personnel or flaws in technology, and underestimated barriers to evaluation could affect negatively the evaluation process and data quality.

Similarly, the results of the comparison indicated that the following two process steps were commonly missing in both guidelines: "Considering methods of data analysis" and "Learning approaches from related projects." The methods of data analysis could be considered while planning an eHealth evaluation, as they can provide insight on the appropriateness and feasibility of the chosen methodological approach. The step "Learning approaches from related projects" can help address the problem of methodological heterogeneity among related studies of eHealth evaluation. Based on the stated benefit of this step in the DECI, supplementing the guidelines for eHealth evaluation planning, with the advice to study and consider using methodological approaches and evaluation standards that were applied in related studies, is necessary. Several other steps that were missing in the AHRQ guideline were related to defining expected results and setting up and monitoring of data collection. The definition of the expected results sets a frame of reference and a benchmark for the outcomes of the study. The steps related to setting up and monitoring of data collection should be considered a part of a feasibility analysis, as the activities included in these steps can reveal that certain outcome measures may be technically unfeasible or overly ambitious in terms of time and human resources needed.

The findings also indicated that there might be circumstances when the GEP-HI or AHRQ guidelines cannot be followed fully. The analyzed case revealed that the steps related to the coordination of the evaluation budget, description of the technical solution, identification of the involved stakeholders, and the phase 4 related to project planning were not performed in the DECI evaluation planning process, because the all these elements were known from the beginning of the project. Also, feasibility analysis in DECI concerned the potential usage of different measures, while the guidelines present the feasibility analysis in relation to the costs and resources.

Finally, the findings demonstrated the potential of assessing adequacy of standards to the practice of eHealth evaluation. This process can help improve the theoretical base and enhance the practice of eHealth evaluation.

4.3 Paper 3: Standards as applied in reality: A case study on the translation of standards in eHealth evaluation practice

The aim of this study was to explore how standards are used in a practical setting of an eHealth evaluation, and to identify the factors that can hinder their use. The following two different types of standards were considered for analysis: 1) an eHealth evaluation framework, namely, the Model for Assessment of Telemedicine Applications (MAST) (Kidholm et al., 2012); and 2) three standardized measures, namely, the EQ-5D-5L for quality of life evaluation (EuroQoL Group, 1990, Janssen et al., 2013b), the Patient Satisfaction Questionnaire (PSQ-18) for patient satisfaction evaluation (Marshall and Hays, 1994), and the Camberwell Assessment of Need for the Elderly-Short Form (CANE-S) for patient needs (Orrell and Geraldine, 2004; Reynolds et al., 2000). The analytical framework of this study combined two perspectives, namely, the process framework of the development of cooperative inter-organizational relationships (Ring and Van de Ven, 1994) and translation theory, including three translation strategies (i.e., *copying*, *addition*, *omission* and *alteration*) (Latour, 1987; Røvik, 2007; Czarniawska and Sevón, 2011).

The analysis of the MAST translation revealed that *addition* strategy was applied when translating it to the DECI context. The factor that led to this strategy was related to the contractual obligations of the DECI. Although the evaluation domains recommended by the MAST were a good fit with the DECI context, the contract of the project included one more domain, that is, "professional satisfaction." Therefore, "professional satisfaction" was *added* in the DECI evaluation framework, next to the domains suggested by the MAST.

The translation of the EQ-5D-5L resulted in *copying* strategy, after the previously considered standard QoL-AD (Logsdon et al., 1999) was ruled out because the latter was not translated and validated in some of the DECI pilot locations. Hence, absence of a validated local language version in a particular location is a factor that can hinder the use of a standard. However, the EQ-5D-5L was used in its original content (*copied*).

The standardized measure PSQ-18 was initially included in the evaluation methodology. However, after it started to be used in the project, the partners noticed that patients with cognitive impairment could not comprehend the questions (a limiting factor). Then, a short interview protocol was created while keeping some elements of the PSQ-18. Hence, the translation of the PSQ-18 employed *alteration* strategy.

The translation of the CANE-S employed *alteration* strategy as well. The hindering factors in the use of the CANE-S in its original content were the limited human resources available to conduct the evaluation of this instrument, and certain partners lacked the experience to use it. Owing to these factors, this quantitative instrument was altered into a customized qualitative interview while retaining some of its elements.

The results of the study indicated that the factors mentioned above can determine the variance

in adherence and create heterogeneity of methodologies between different eHealth evaluation studies. Several solutions to cope with heterogeneity were identified. First, the standards need to be translated and validated in different contexts. Second, eHealth evaluation studies are often demanding and include a set of different standards and evaluation activities. Therefore, at times, a standard can be translated using the omission or alteration strategies because of the lack of human or financial resources dedicated to the entire evaluation assignment. Moreover, there is a higher probability that a standard will be used if the evaluators already have an experience in applying such standard.

The study also demonstrated that the use of standards is determined by the collaborative activities and uncertainties concerning different partners in the research consortium. A decision leading to a certain translation strategy is reached when the partners bargain and resolve these uncertainties collaboratively to reach a consensus.

5   DISCUSSION

The purpose of this thesis was to study the use of standards in eHealth evaluation practice, and provide insight on theory and practice of eHealth evaluation. The three papers covered different types of standards and goals. The first paper focused on investigating the actual use of conceptual eHealth evaluation frameworks in practice. The second paper investigated the fit of two eHealth evaluation planning guidelines with practice. The third paper explored the factors affecting the (non)use of different standards, such as an evaluation framework and several standardized metrics, in eHealth evaluation practice. This chapter aims to integrate the findings from all three studies and discuss them in relation to previous research.

5.1   Explaining the (non)use of standards in eHealth evaluation practice

5.1.1   The actual use of standards in eHealth evaluation practice

With the help of Research Question 1 (i.e., "To what extent are eHealth evaluation frameworks used in practice?"), the research findings in this thesis indicated that the use of eHealth evaluation frameworks in empirical evaluation studies is quite low, which supports the concern in the eHealth evaluation literature regarding the limited use of standards (Greenhalgh et al., 2004; Ammenwerth, 2004; Proudfoot et al., 2011; Janssen et al., 2013a; Mookherji et al., 2015). The research concerning Research Question 1 was limited to eHealth evaluation frameworks and did not focus on other types of standards, such as eHealth evaluation planning or reporting guidelines or standardized metrics. Additionally, the reasons of not referring to the conceptual eHealth evaluation frameworks in empirical studies remained unclear. Since many eHealth evaluation studies might not be published, the assessment of the actual usage of standards is problematic. However, an additional review of the published literature on the actual usage of other types of standards can provide a more comprehensive picture.

5.1.2   Adequacy of standards for eHealth evaluation practice

With the help of Research Question 2 (i.e., "To what extent are standards adequate in a practical eHealth evaluation setting, and how can standards be improved?"), the research indicated that the analyzed standards were useful and practical to various degrees. Findings from study 1 demonstrated that the analyzed eHealth evaluation frameworks would have been adequate for the research questions in many of the reviewed empirical studies. With the help of evaluation frameworks, the empirical studies could have improved quality and credibility. By not using the evaluation frameworks, the scholars risk missing to address the evaluation domains defined as a good practice for essential evidence generation (Mookherji et al., 2015).

Findings from study 2 also indicated that the eHealth evaluation planning guidelines, namely, the Health Information Technology Evaluation Toolkit (AHRQ) and Guideline for Good Evaluation practice in Health Informatics (GEP-HI) are quite adequate for a planning process

in a practical setting but can be improved further. From the identified "matches" between the guidelines and practical activities surrounding the eHealth evaluation planning, the research conducted showed that the guidelines can be a sound support to evaluation practitioners offering a step-by-step guidance for evaluation planning.

Several areas of improvement of the eHealth evaluation planning guidelines were identified to enhance their adequacy for practice. Overall, the prerequisites of quality in evaluation studies (Mookherji et al., 2015) and the lack of it are determined by the decisions taken in the planning phase. Hence, eHealth evaluation guidelines could be considered as a way to reduce the number of evaluation quality issues related to scientific rigor, which has been raised by numerous scholars (Proudfoot et al., 2011; Mookherji et al., 2015; Cowie et al., 2016). Specifically, among the suggestions is that the AHRQ and GEP-HI guidelines could be updated with a recommendation to consider the methodologies used in other evaluation studies at the beginning of the evaluation planning. After considering the methodologies used in other studies, evaluators can choose a methodology for their particular situation. Hence, the present thesis showed that standards can be a good means to address problematic areas in eHealth evaluation, such as the lack of methodological uniformity or scientific rigor (Ammenwerth, 2004; Greenhalgh et al., 2004; Proudfoot et al., 2011; Janssen et al., 2013a; Mookherji et al., 2015; Cowie et al., 2016).

5.1.3 *Factors limiting the use of standards in eHealth evaluation practice*

A model described in Figure 13 demonstrates how "evaluation reality," enacted through interorganizational collaboration (Ring and Van de Ven, 1994) and translation processes (Latour, 1987; Røvik, 2007) within a particular context, shapes the use of standards in eHealth evaluation practice. Figure 13 is a development of the Figure 4. It includes a range of adherence and the factors identified from the research conducted.
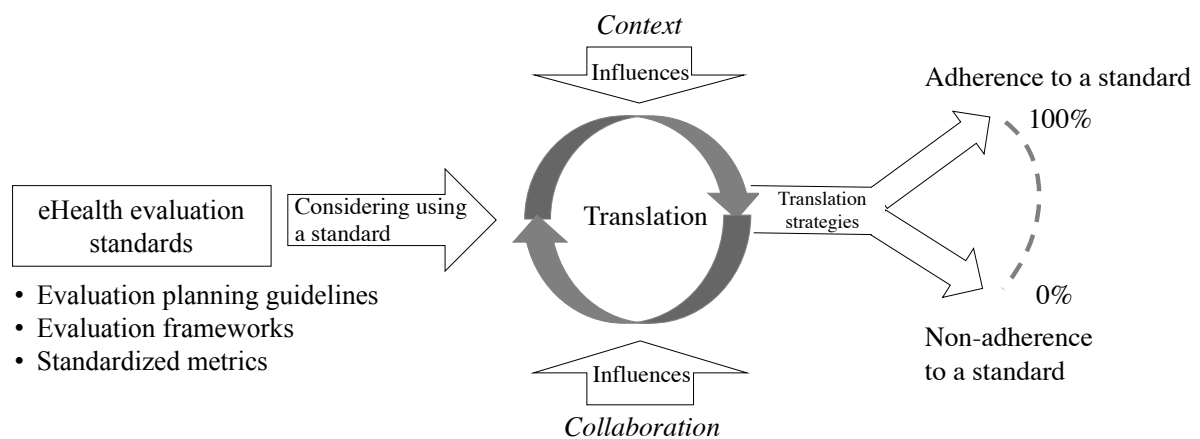
Figure 13. Model of translation of standards in summative eHealth evaluations

The present thesis demonstrated that when an eHealth evaluation standard (a guideline, a framework, or a standardized metric) is considered to be used in an evaluation study, it enters a translation process before its use in a study is decided. Translation does not happen in an isolated environment and "reality" (i.e., the context in which a study takes place and people perform processes and activities to plan an evaluation) plays a role. This thesis showed that the contextual aspects can include a geographical location, a research set-up of an evaluation study, or terms and conditions defined in the contract of a study. These elements can cause barriers to the use of standards, as shown in studies 2 and 3. Moreover, in a multiparty research, collaborative activities and involvement of multiple contexts can create additional complexity. The barriers emerge when several parties from different disciplines and contexts have to reach consensus upon the common standards to be used during the evaluation. The findings support previous research, claiming that evaluations can be challenging owing to a complex social reality (Stufflebeam, 2001; Bates and Wright, 2009). Hence, contextual factors and interorganizational collaboration shape the results of translation. The resulting (i.e., translation strategies [Røvik, 2007]) lead to different outcomes in terms of adherence to standards. The present thesis demonstrated that there can be a range of outcomes in terms of adherence to standards in eHealth evaluation practice. Moreover, the effects of different translation strategies on adherence to standards are explained.

Translation strategies can have different effects on the use of standards in eHealth evaluation. *Copying* strategy does not change the content, shape, or methodological approach of a standard,

and creates conditions for methodological uniformity among different studies. It also enhances credibility and facilitates publication of the results of a study. However, a *copied* standard might not be a perfect fit with a specific context of an evaluation study (for example, it can be generic, as demonstrated in study 3). *Addition*, *omission*, and *alteration* strategies change a standard's content, shape, or methodological approach, but in such cases, a standard might fit with a context of a study better. However, these three strategies can hinder comparability of study results among studies and thus the publishing of the results of evaluation.

Findings of this thesis suggested that translation strategies (Røvik, 2007) may have different implications regarding the use of standards, depending on a type of a standard. eHealth evaluation planning guidelines and evaluation frameworks provide guidance to evaluation practitioners and are expected to be adapted in a particular context. Meaning, it is acceptable to assume adherence to these standards if they supported evaluation planning activities in a certain way (partial adherence). Hence, all translation strategies imply adherence to the guidelines and frameworks, where omission and alteration strategies indicate lesser adherence (because certain elements would be removed or changed). As suggested in study 2, unawareness of the existence of guidelines can be a factor limiting the latter's use in evaluation studies. Study 1 also indicated that the lack of awareness of the existence of a particular evaluation framework could be attributed to the finding that there was almost no reference to the frameworks in the set of scientific publications studied. Additionally, the frameworks might not be used in practice when their recommended evaluation questions (evaluation domains) do not fit the scope of the evaluation study.

The purpose of standardized metrics in evaluation is to provide a quantitative result. Every element of a standard has a role in the calculation of the score. Therefore, copying strategy implies adherence. Other translation strategies can also be used for standardized metrics to better suit a particular study, but then adherence to a standard cannot be implied, as the score calculation would not be applicable. As suggested in study 3, factors limiting adherence to the standardized metrics are a standard's inadequacy to address a patient group or disease and non-validity in a particular language or location. Non-adherence to a standard may also occur when users of a standard possess insufficient experience or resources to use it.

Table 2 provides a summary of the factors hindering the use of standards by different standard types studied. The factors have been derived from the three studies conducted.

Table 2: Summary of factors hindering the use of standards in eHealth evaluation

Type of eHealth evaluation standards	Degree of adherence	Translation strategies (Røvik, 2007) that indicate adherence	Factors that can limit the use of or hinder adherence to standards in practice
Evaluation planning guidelines (e.g., GEP-HI, AHRQ)	Different degrees of adherence are appropriate.	Copying, Addition, Omission, Alteration	Evaluator's unawareness of the guidelines
Evaluation frameworks			Evaluator's unawareness of a framework Does not fit a study's scope
Standardized metrics	Either adherence or non-adherence	Copying	Inadequacy to address a target population or a disease Non-existence of a validated version of a standard in a particular location Insufficient experience to use a standard Insufficient resources to use a standard

To sum up, this thesis addresses the knowledge gap on the actual use of standards in eHealth evaluation practice, by identifying factors leading to variations in adherence to standards in a practical setting of an eHealth evaluation. Previous research has identified challenges when it comes to evaluation in a multiparty research. The previously identified challenges were the alignment of goals, resources, and capabilities, as indicated by Greenhalgh and Russel (2010). The research described in this thesis has identified that resources and capabilities are the factors that can also hinder the use of standards in a multiparty research collaboration. Besides that, this research has identified additional factors (see Table 2) that can affect the use of standards in eHealth evaluation. With the growing funding for multinational and multidisciplinary research, such challenges will be increasingly affecting eHealth evaluation studies.

5.2 Implications for theory and practice

The following chapter summarizes the key implications for theory and practice developed from the conducted research.

5.2.1 Implications for theory

This thesis suggests that to achieve quality in an eHealth evaluation study, a balance between standards and evaluation reality needs to be established. When standards are adapted or translated in a particular context, the quality of evaluation improves because standards are adapted in a way that they are relevant to the contextual circumstances (e.g., purpose and scope of evaluation defined in a contract, specificities of evaluation subjects, capabilities of evaluators, and resources dedicated to a study). Similarly, eHealth evaluation practice benefits from standards by providing credibility and creating conditions for comparability of results across different studies. The interplay between standards and reality is an additional view on quality in eHealth evaluation. Previously, quality in eHealth evaluation has been understood in relation to study design and 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 (Mookherji et al., 2015). The idea that trade-offs between standards and reality occur in eHealth evaluation practice needs to be embraced when addressing the issue on the insufficient use of standards in eHealth evaluation studies.

In addition, this thesis suggests that adherence to standards in eHealth evaluation can be understood as a range that depends on the type of a standard. In previous research, adherence to standards has been displayed as a scale, that is, a study either adheres to a standard or not. Scholarly discussions around this matter have not been extensive; adherence has only been mentioned as necessary (e.g., Mookherji et al., 2015; Cowie et al., 2016) and lacking in practice (Greenhalgh et al., 2004; Ammenwerth, 2004; Proudfoot et al., 2011; Janssen et al., 2013a; Mookherji et al., 2015).

Finally, this thesis contributes to the field of eHealth evaluation by combining the concept of a standard with the translation theory (Latour, 1987; Røvik, 2007; Czarniawska and Sevón, 2011). Such combination leads to the creation of a model of translation of standards in summative eHealth evaluation practice (Figure 13).

5.2.2 Implications for practice

The findings in this thesis lead to several practical propositions for eHealth evaluation practitioners:

(1) The eHealth evaluation planning guidelines (e.g., the AHRQ and GEP-HI) could be updated for better alignment with practice. Specifically, several steps could be added with regard to: 1) learning approaches from related projects, 2) considering methods of

data analysis, 3) defining expected results, 4) setting up data collection, and 5) setting up monitoring of data collection.

(2) When using standards in eHealth evaluation practice, practitioners should be explicit if, due to practical circumstances, certain elements of the standard cannot be addressed. The deviations need to be documented in evaluation reports. Such approach is expected to enhance research credibility.

(3) When planning an eHealth evaluation, practitioners and society would benefit from investigating the methodological approaches and standards used in other studies through a literature review or by contacting the evaluators. Such approach is expected to increase opportunities for methodological uniformity across different studies.

(4) eHealth evaluation planning guidelines should be used by practitioners, as they can enrich and guide the planning process and enhance quality of the evaluation study.

(5) Practitioners should not underestimate the social complexity that emerges during an eHealth evaluation planning, especially in a multinational and multidisciplinary study. Extra time may be required to resolve challenges emerging from social complexity.

(6) As the eHealth evaluation community recognizes the importance of standards and has asked for guidance, supporting practitioners by publishing methodologies adequate for practice is deemed necessary (Mookherji et al., 2015; Cowie et al., 2016). Therefore, it is valuable to continuously investigate the adequacy of different standards for practice.

(7) The standards should include guidance to practitioners on how to select and use standards, to maintain methodological uniformity among similar studies.

6 CONCLUSIONS, LIMITATIONS, AND FUTURE RESEARCH

6.1 Conclusions

This thesis aimed to study the use of standards in eHealth evaluation practice, and provide insight on theory and practice of eHealth evaluation. The key contribution of this thesis is the identification of a range of adherence to standards in eHealth evaluation practice and of reasons for this phenomenon. Because eHealth solutions are implemented in various contexts under many different circumstances, the standards are contextualized using different translation strategies (Røvik, 2007). To arrive at a decision to use a particular translation strategy, it may take rounds of negotiations between involved parties (Ring and Van de Ven, 1994), thus creating additional complexity that hinders the use of standards. Hence, a trade-off between standards and context is required to achieve relevance of the standards to a particular study and a high-level quality of evaluation.

Based on the three conducted studies, an actionable advice for eHealth evaluation scholars and practitioners is offered to improve the uptake of standards. To tackle standardization and challenges concerning adherence to standards in eHealth evaluation practice, scholars could benefit from viewing the use of standards as a dynamic process of translation (Røvik, 2007), with multiple possible outcomes, leading to a particular level of (non)adherence to a standard. Moreover, scholars could benefit from investigating the reasons for (non)adherence to standards in different contexts and from various perspectives (e.g., the (non)adopter of a standard). Additionally, existing standards could be improved to better serve evaluation practitioners. In turn, practitioners could benefit from the use of standards, and report deviations leading to a lesser or non-adherence to them, when conducting an eHealth evaluation research. In this goal, practitioners should be supported by the standards that provide guidance in selection and use of standards, to create better chances for methodological uniformity among similar studies.

The combination of standards in eHealth evaluation and the translation theory may help scholars facilitate understanding of the complexities surrounding the use of standards in eHealth evaluation. This thesis enabled the identification of the reasons standards were used insufficiently in a specific case on eHealth evaluation.

6.2 Limitations

The research described in this thesis has several limitations. In study 1, only eHealth evaluation frameworks as a type of standards were researched. The actual usage of other types of standards

was not explored. The implication of this limitation is related to the established fact that standards are used insufficiently in eHealth evaluation studies. An investigation of the actual use of other types of standards may provide a better picture of the problem areas.

Another limitation is related to the case study design in studies 2 and 3. This limitation concerns the theoretical generalization of the research findings. Moreover, data for studies 2 and 3 partly overlapped but served different purposes in both studies.

6.3   Future research

This chapter describes potential pathways for future research.

Based on the findings in this thesis, it would be interesting to explore the "voice" of eHealth evaluation practitioners as potential adopters of standards. The potentially interesting questions of research could be: Why would different evaluation practitioners choose to adopt the same standard, thus creating conditions for methodological uniformity among themselves? What determines the selection of a standard in comparison with alternative ones that serve similar purposes? As previously noted, uniformity created by standards can imply that certain adopters give away their own freedom to choose a standard for the benefit of others (Brunnson and Jacobsson, 2000). An international survey can be conducted among evaluators of eHealth to obtain insight on these research questions. Such research is expected to determine effective mechanisms that can enhance the uptake of standards in eHealth evaluation practice.

Another pathway for research opens after the interventions of the Digital Environment for Cognitive Inclusion (DECI) project has been completed in Sweden, Italy, Spain, and Israel, and the data has been collected. There is a potential to elaborate the data collected in different evaluation domains, such as patients' and health care professionals' satisfaction, technology acceptance, quality of life, and adherence to treatment. Future research can aim to determine causal relationships between different indicators in the DECI dataset and try to understand the interactions among those indicators with respect to outcome variables. Such research could have several possible benefits. First, knowledge of how different indicators influence each other could help decide what features of eHealth solutions could be included and improved, and to which features investments could be directed to maximize the benefits for patients and health care systems. Moreover, the uncovered causal relations will allow to validate the existing limited theory on relations among different indicators concerning eHealth solutions for the elderly suffering from mild cognitive impairment and mild dementia. This validation will pave new ways in improving the quality of health care provisions for these populations through more informed and targeted designs and evaluations of eHealth programs.

REFERENCES

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.

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.

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.

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

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.

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.

Burghgraeve, P., & De Maeseneer, J. (1995). Improved methods for assessing information technology in primary health care and an example from telemedicine. *Journal of Telemedicine and Telecare*, *1*(3), 157-164.

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.

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.

Cunliffe, A. L., & Karunanayake, G. (2013). Working within hyphen-spaces in ethnographic research: Implications for research identities and practice. *Organizational Research Methods*, *16*(3), 364-392.

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.

Dekker, H. C. (2004). Control of inter-organizational relationships: evidence on appropriation concerns and coordination requirements. *Accounting, Organizations and Society*, *29*(1), 27-49.

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.

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.

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.

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.

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

Eysenbach, G., & Consort-EHEALTH Group. (2011). CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. *Journal of Medical Internet Research*, *13*(4).

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

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

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.

Global diffusion of eHealth: making universal health coverage achievable. Report of the third global survey on eHealth. Geneva: World Health Organization; 2016. License: CC BY-NC-SA 3.0 IGO.

Grbich, C. (2012). *Qualitative Data Analysis: An Introduction*. London, UK: Sage.

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.

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.

Halldorsson, A., & Aastrup, J. (2003). Quality criteria for qualitative inquiries in logistics. *European Journal of Operational Research*, *144*(2), 321-332.

Han, Y. Y., Carcillo, J. A., Venkataraman, S. T., Clark, R. S., Watson, R. S., Nguyen, T. C.,

Hülya, B., & Orr, R. A. (2005). Unexpected increased mortality after implementation of a commercially sold computerized physician order entry system. *Pediatrics*, *116*(6), 1506-1512.

Hart, C. (1998). *Doing a literature review*. London, UK: SAGE.

Hebert, M. (2001). Telehealth success: evaluation framework development. *Medinfo*, *10*, 1145-9.

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.

Hornby, A. S., Cowie, A. P., & Lewis, J. W. (1974). adequacy. *Oxford advanced learner's dictionary of current English*. London: Oxford University Press. Retrieved from https://www.oxfordlearnersdictionaries.com/definition/english/adequacy

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.

Janssen, M. F., Pickard, A. S., Golicki, D., Gudex, C., Niewada, M., Scalone, L., Swinburn P., & Busschbach, J. (2013b). Measurement properties of the EQ-5D-5L compared to the EQ-5D-3L across eight patient groups: a multi-country study. *Quality of Life Research*, *22*(7), 1717-1727.

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.

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.

Koppel, R., Metlay, J. P., Cohen, A., Abaluck, B., Localio, A. R., Kimmel, S. E., & Strom, B. L. (2005). Role of computerized physician order entry systems in facilitating medication errors. *JAMA*, *293*(10), 1197-1203.

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.

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

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.

Logsdon, R. G., Gibbons, L. E., McCurry, S. M., & Teri, L. (1999). Quality of life in Alzheimer's disease: patient and caregiver reports. *Journal of Mental health and Aging*, *5*, 21-32.

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).

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.

Merton, R. K. (1987). Three fragments from a sociologist's notebooks: Establishing the phenomenon, specified ignorance, and strategic research materials. *Annual Review of Sociology*, *13*(1), 1-29.

Moher, D., Weeks, L., Ocampo, M., Seely, D., Sampson, M., Altman, D. G., Schulz, K. F., Miller, D., Simera, I., Grimshaw, J. & Hoey, J. (2011). Describing reporting guidelines for health research: a systematic review. *Journal of Clinical Epidemiology*, *64*(7), 718-742.

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.

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

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.

Oh, H., Rizo, C., Enkin, M., & Jadad, A. (2005). What is eHealth (3): a systematic review of published definitions. *Journal of Medical Internet Research*, *7*(1).

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).

Pittaway, L., Robertson, M., Munir, K., Denyer, D., & Neely, A. (2004). Networking and innovation: a systematic review of the evidence. *International Journal of Management Reviews*, *5*(3-4), 137-168.

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.

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.

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.

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

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

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.

Savedoff, W. D., Levine, R., Birdsall, N. (2006). *Report of the Evaluation Gap Working Group*. Washington, D.C.

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).

Sockolow, P. S., Crawford, P. R., & Lehmann, H. P. (2012). Health services research evaluation principles. *Methods of Information in Medicine, 51*(02), 122-130.

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).

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.

Tranfield, D., Denyer, D., & Smart, P. (2003). Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *British Journal of Management*, *14*(3), 207-222.

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

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).

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.

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

WHO, 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.

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

Yin, R.K. (1989). *Case study research*. Newbury Park, CA: SAGE Publications.

Yin, R.K. (2015). *Case study research. Design and Methods*. London, UK: SAGE Publications.

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.