Universiteit Utrecht

MASTER'S THESIS

Data Quality of the Dutch DBC Information System

P.C.E. van Stijn

3019985

November 26, 2012

Version 1.0

Public Version

Supervisor Deloitte:
Martijn Ludwig

Deloitte.

Supervisors UU:
Ronald Batenburg
Marco Spruit

Institute of Information & Computing Sciences
ICS

Deloitte.

Name:	Peter Carlo Ellert van Stijn
Student number:	3019985
Contact:	pcevanstijn@gmail.com

Title thesis:	Data Quality of the Dutch DBC Information System
Date:	November 26, 2012

University:	Utrecht University
Faculty:	Information & Computing Sciences
Study:	Business Informatics
Company:	Deloite
Supervisors UU:	Ronald Batenburg
	Marco Spruit
Supervisor Deloitte:	Martijn Ludwig

Deloitte.

Contents

Part I

Problem Statement & Background

1 Introduction

In the Netherlands, hospitals are reimbursed based on fixed prices for a combination of diagnosis and treatment: a DBC (Dutch: *Diagnose Behandeling Combinatie*). This hospital funding model has been introduced in January 2005. One of the most important drivers of the introduction of the system was to increase transparency on price and performance of the Dutch (hospital) healthcare [22, 40]. Furthermore, managed competition between health care providers and health insurers was introduced with the DBC system. Health insurers are (to a certain extent) free to decide which care products from which hospitals will be purchased. For these care products, hospitals only receive reimbursements if the client is attached to one of their contracted insurers.

Efficiency and Quality of care incentives from the DBC model This managed competition provides the hospitals incentives to improve the quality of care while simultaneously improving the efficiency [22, 38]. If the quality of care is not guaranteed or insufficient, insurers can refrain from purchasing the care products. Furthermore, hospitals receive a fixed reimbursement for a delivered diagnosis and treatment combination (DBC) and insurers negotiate for minimizing these prices. Therefore, the hospitals will be inclined to improve efficiency by minimizing inessential activities. Examples of this are reduction of duplications, unnecessary tests and reducing the length of stay [19].

This also implies that with the DBC system, the concepts of healthcare quality and efficiency are becoming increasingly related. Due to earlier described quality incentives of the DBC system, an increased efficiency should not negatively influence the quality of the provided care. An important dimension and indicator of the (relatively difficult to measure) concept of quality of care is patient safety [55]. It refers to a situation in which a patient has no (or little chance of) the occurrence of an incident [83]. Thus, although there are incentives to improve the efficiency of the provided care, patient safety should be guaranteed. When analyzing health care performances, these two dimensions should therefore ideally be combined to ensure this relation is taken into account.

The DBC Information System The DBC system is accompanied by an extensive registration, which is gathered in a national database: the DIS (DBC Information System) [21]. This database includes all data relevant to the hospitals' declarations and to the maintenance of the reimbursement system, structured in a standardized way. It includes all diagnoses and activities performed by all hospitals on all patients in the Netherlands. Furthermore, it contains some basic information on the patients (e.g., unique identifiers, gender and year of birth) and information on the derived care product which was reimbursed by the insurer. The organization involved with the database is also referred to as DIS and is responsible for the gathering, validation and storing, as well as the distribution of these data for further analysis.

1.1 Problem Statement

Data understanding is an important precondition for data analysis (e.g., the Crisp-DM cycle [89]). Assessment of the data quality is part of this phase, as bad data quality negatively influences the reliability of the results of analyses on those data [7]. This implies it is essential to assess the DIS data quality before performing analyses. However, this is not described in publications or official reports yet.

Besides the intrinsic quality of the data itself, part of the perceived quality depends on the *"fitness for use"* [74]. This implies that when assessing data quality, this fitness for use, or usefulness, in a specific context should be taken into account. This aspect of data quality is also referred to as contextual data quality [57, 85].

When the national DIS database was introduced, some other national databases were phased out to be replaced by the DIS. First of all, the LMR (Dutch: *Landelijke Medische Registratie*) was no longer mandatory for hospitals to provide. Another database, the EJ (Dutch: *Enquête Jaarcijfers*) was greatly reduced in terms of included data [27]. However, these national databases had been used for research in the contexts of Dutch patient safety (LMR) and efficiency (EJ) for years. Consequently, after phasing out of these databases, no new Dutch hospital efficiency studies have been performed, at least not on current data. The patient safety researches found limitations due to reduced quality of the used LMR database. The latter problem was encountered in a recently published report on hospital death numbers (HSMR): some hospitals were entirely missing or their results were not plausible.

Currently, the DIS dataset is mainly used for maintenance of the reimbursement system [22], and for some high-level analyses on Dutch healthcare [27]. With the reduction of the quality of the other national databases and the incentives intrinsic to the DBC system, it will be interesting to assess whether the DIS database is also useful for both efficiency and patient safety researches. This is especially the case as the DIS was aimed to replace the two systems that historically analyzed these factors in the first place. However, the quality in these contexts have not been assessed yet either.

Thus, the intrinsic data quality of the DIS data has not been examined in scientific research yet, although knowledge on this quality is an important precondition for data analyses. Furthermore, the contextual data quality of these data in the fields of efficiency and patient safety research have not been assessed yet, although these are relevant and related contexts and older methods to measure these aspects are no longer (satisfyingly) possible.

1.2 Research Statement and scope

The main goal of this research is to assess both this intrinsic and contextual data quality of the DIS database. The main research question is therefore formulated as follows:

> RQ: ***What is the intrinsic and contextual data quality of the DBC Information System within the Dutch hospital care, and how are these determined and interrelated?***

The DBC system is not only used in the Dutch hospital care, but also in the other cure sectors such as the mental health (Dutch: *Geestelijke Gezondheidszorg*). This thesis focuses on the DIS data quality and related organizational issues only in hospital healthcare.

Currently, the DIS data are distributed to a number of organizations legally entitled to these data such as the NZA, CBS and DBC Maintenance. Furthermore, the data are distributed to organizations that have explicit permission from the concerned healthcare providers, currently only Dutch Hospital Data (DHD) [27]. Analyses on the data can therefore currently only be performed by these organizations. Furthermore, there is no reference material available yet to compare the data with. Therefore, quantitative approaches to assess the intrinsic data quality are not possible in this research context.

Consequently, this research is scoped to a qualitative (process-based) measurement. With this approach, the organizational and technical issues that (negatively) influence the data quality are identified and evaluated in order to approximate the intrinsic quality. The first sub question is on this part of the research and is formulated as follows:

> sQ 1: *What is the intrinsic data quality of DIS and how is this determined by organizational and technical issues?*

The last two sub questions pertain to the second part of the research question, the contextual quality. Two contexts to evaluate this in have been elected: efficiency and patient safety.

> sQ 2: *What is the contextual data quality of DIS to evaluate efficiency in Dutch hospitals and how is this determined by intrinsic data quality of DIS?*
> sQ 3: *What is the contextual data quality of DIS to evaluate patient safety (underlying quality of care) in Dutch hospitals and how is this determined by intrinsic data quality of DIS?*

The contextual data quality is assessed as the ”fitness for use” (usefulness) [74, 85] for efficiency and patient safety evaluation. Naturally, many other contexts could be examined. However, the scope is on these two because of their relation to the system, as described earlier. Furthermore, it is to determine whether this database could replace the databases that were earlier used for efficiency and safety researches.

1.2.1 Conceptual Model

The relation between the research questions and the concepts related to data quality is shown in Figure 1. As can be seen, the main research question concerns the overall data quality of the DBC Information System, which contains the quality intrinsic to the data, and the quality related to the context of use. As explained, the intrinsic quality can either be measured quantitatively (P1), or approximated qualitatively (sub Question 1). Proposition 2 (P2) states that ”intrinsic data quality influences the contextual data quality”, i.e., when the intrinsic data quality is insufficient, the contextual quality will be lower as well. This also implies that the organizational issues that are identified and influence the intrinsic quality will ultimately also influence the ”fit for use” in a specific context. The two contexts for which the data quality is examined in this research is efficiency (sQ2) and patient safety (sQ3). P3 states that ”patient safety is a sub dimension of quality of care” [55].

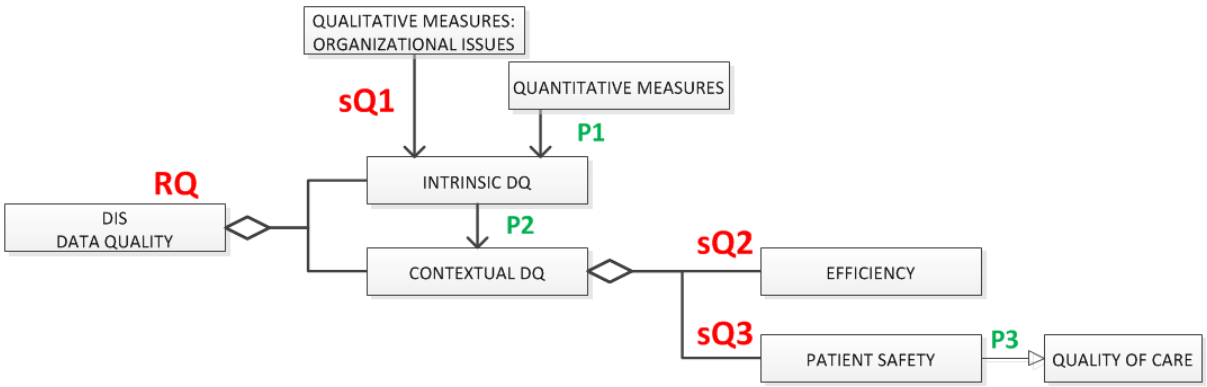

Figure 1: Conceptual model explaining the relation between the sub questions and used concepts

1.3 Scientifc and Practical Contributions

Practical Contributions The DIS data is currently already used for some analysis, especially for the maintenance of the DOT system (e.g., creating new care products) and for the calculation of cost prizes. By approximating the intrinsic DIS data quality through identification of the organizational and technical issues that influence this, more understanding on the reliability of these current analyses will be created. Furthermore, more insight in the registration behavior within the hospitals will be created. With knowledge on the organization issues that are currently present and influence the data quality, process improvements to improve this quality can be initiated. This thesis will also provide recommendations for these improvements based on the findings. With it, the registration processes, the issues intrinsic to the DBC and DOT system and eventually the DIS data quality can be improved, as well as possibilities of these data for analyses – especially on efficiency and patient safety research. Furthermore, there is currently a public debate on efficiency versus quality of care and whether health insurers are (or should be) able to determine the efficiency and quality performances of hospitals. By looking at the possibilities to measure both efficiency and safety from the same database, and thus creating comparable results, this thesis research aims to add to this discussion.

Scientific Contributions Besides these practical contributions, there are some additions to the scientific knowledge base. First of all, composing a useful data quality (measurement) framework to measure the level of quality of data is the starting point of any data quality related activity [6, 7]. This thesis assembles different ”standard” quality frameworks to one deemed useful for this situation and applies it to this healthcare database. A scientific contribution of this thesis in that sense is the construction and applying (testing) of a data quality framework specifically aimed at healthcare related research.

Furthermore, this research aims to add to the current research on hospital efficiency and patient safety. First of all, it assesses the usefulness of the DIS database to evaluate both aspects. This is necessary as databases used for earlier research are no longer reliable or suitable, as explained earlier, and new possibilities have to be examined. Second, the research summarizes the current assessment methods for efficiency and safety used both in scientific researches and in practice (internal to the hospital). Besides essential for the contextual quality analysis, it will create understanding of the different approaches which could bring the theory and practice closer. Third, the research evaluates the possibilities to extract indicators pointing at the occurrence of incidents (patient safety issues) from the operational DIS data. This is not yet performed in the Netherlands, but would add to the current researches and could potentially reduce the currently manual and time consuming safety analyses.

1.4 Thesis Structure

This section has provided the problem statement, research statement and stated the research questions of this thesis. The following Section will then describe more on the approach used to answer these research questions: a qualitative approach of performing expert interviews. Section 3 provides background information on the Dutch DBC hospital reimbursement system and the DBC information system (DIS). After this, Section 4 provides a scientific basis of the research, separated in three subsections covering the three main aspects of this research: data quality, efficiency and patient safety. This concludes the first part of the thesis, which provides all necessary background on the research and research topic.

The second part of the thesis describes the results and analysis. Section 5 first describes the interview results. For this Section, two versions exist: a confidential version which is only received by the supervisors and which contains full information on the interviewed experts. The second version is the ”public” version, where no identifying information on the experts is provided. The analysis of the interview results is performed in four phases. First, the DIS data delivery process is fully analyzed based on the interviews, as this is an important input for the rest of the analysis. This is described in Section 6. Section 7 then describes the analysis of the intrinsic data quality; Sections 8 and 9 consecutively describe the analysis of the contextual quality for efficiency and patient safety research.

The final part of this Thesis is started with a conclusion section answering the research questions (Section 10). Section 11 provides recommendations for improving certain organizational and technical issues identified during the research. The thesis is finished with a discussion section in Section 12, reflecting the results, research method and providing future research directions.

2 Research Method

In order to answer the sub questions and ultimately the research question, a qualitative approach of performing expert interviews is followed. The reason for this, the sampling of the experts, the structuring of the results and further information on this process is described in Section 2.2. Before this empirical part of the research, desk research is performed to provide a theoretic foundation of the topic and prepare for the interviews. This includes a literature research, defining and categorizing the used concepts, and further background information gathering and is described in Section 2.1. The steps performed for the analysis of the gathered data follow the *Grounded Theory* approach [24] and ultimately leads to the conclusions and recommendations. Section 2.3 provides a description of this analysis process. The section concludes with a discussion on validity issue safeguards in Section 2.4.

2.1 Desk Research

2.1.1 Background Research on the DBC system and DIS

Background knowledge on the Dutch DBC (and DOT) system and the associated DIS database is one of the first requirements to this research. The information is mainly gathered from technical reports, although a limited amount of scientific publications exist as well. Regarding the DIS database, this research phase focuses mainly on the data contents (e.g., the entity-relationship model). Section 3 describes the results of this research phase. The organizational processes associated to the creation of the DIS data were mainly identified through the expert interviews: the results of this research phase are part of the analysis and are described in Section 6.

2.1.2 Data Quality research and framework creation

Data quality is a multi-dimensional concept, for which many different dimensions and categorizations have been constructed. Furthermore, there is no general consensus on the exact meaning of these dimensions. Therefore, any data quality related activity should start with choosing and defining the different quality dimensions that will be used to measure the level of quality of data [6, 7, 65]. Different data quality frameworks and methodologies are examined, again with a focus (where possible) on a healthcare setting. This is executed by performing a non-structured literature review. Although a structured approach such as a Structured Literature Review (SLR) [64] would ensure *all* (relevant) frameworks and methodologies are found, this approach has not been chosen as this is not the focus of this thesis and such research has already been performed to a certain extent (e.g., [6, 7]). The results of this background research, with the most important frameworks found, are provided in Section 4.1.1. A combination of the most influential frameworks from literature that include the aspects and definitions relevant to this research is concluded from that research phase (Section 4.1.3). Clear definitions for the dimensions were adopted from the most influential literature sources. The assembled data quality framework is used as main input for the interview outline related to the data quality assessment.

2.1.3 Definitions and categorization of efficiency and patient safety

One of the validity threats described by Yin [92] is construct validity (see Section 2.4), which refers to the extent in which the used concepts are operationalized and measured correctly. Before examining the

possibilities to evaluate hospital efficiency and patient safety based on the DIS data, sufficient definitions of these two and their related concepts have to be formulated.

Most of the definitions on efficiency and patient safety are already provided in literature, other definitions are formulated in an inconsistent way. The aim of this part of the research is to identify the different concepts related to efficiency and patient safety, based on scientific publications. Definitions and categorizations for the concepts are then adapted and combined from different (influential) literature sources. The categorizations will provide useful to describe the (im)possibilities of the DIS data later in the thesis.

Systematic Literature Review An earlier focus of this research has been on solely on incidents in healthcare (patient safety relates to the occurrence of incidents). For this earlier focus, a systematic literature review (SLR) has been performed. Although this review has not been used for this thesis further, the categorization of different types of incidents has been constructed from it (Section 4.3.1).

For this SLR, the first 500 PubMed search results of the search terms ”Incidents” AND ”Analysis” were analyzed (from a total of 12.236, performed between November 24 and December 6 2011). This resulted in 25 relevant papers based on title, of which 9 were deemed useful based on abstract and conclusion and were thus included. Other references for the definitions were found through these papers. For the definitions and categorizations of concepts related to (hospital) efficiency, a ”normal” unstructured literature research is performed.

Further Literature Background Research Besides providing concept definitions and categorizations, also a ”regular” literature background research on the topics of data quality, efficiency and patient safety is performed. For the data quality and efficiency, two research areas not necessarily related to healthcare, the uses and research in a healthcare setting are also discussed. Furthermore, for all topics a description of existing literature with a specific focus on The Netherlands will be provided.

2.2 Empirical Research - Expert Interviews

Reasoning behind the research approach There are several reasons why the qualitative approach of expert interviews is useful and suitable to answer the research questions in this thesis. These are further explained below and relate to impossibility of a quantitative approach, low amount of people knowledgeable with the entire research topic, and other qualitative approach being less suited. Information on data gathering approaches is taken from [4, 13, 34, 37]:

1. Quantitative approaches (tests) are in this case not suitable. Most importantly, no access to the DIS data is granted for a research such as this one. Furthermore, if this would be the case, there is no real ”reference data” this dataset can be measured against. At this moment, the similarities between what the hospitals deliver and what data is included in the dataset, is so unclear the hospitals have already a difficulty to recognize their own data if they get receive DIS data for comparison. This makes quantitative approaches to test for (population) completeness impossible. Furthermore, tests on the column completeness of the data (empty values) are already performed at the validation before the data are included in the DIS database, and invalid data is not included. So these kinds of (quantitative) tests will not provide interesting results. The same problems occur

when trying to assess other dimensions of data quality: the tests are either already performed at validation, or are impossible as a reference set cannot be assessed. It could also be the case the ”real world state” the data is supposed to express cannot be known at all (traced back to the performed care).

2. Surveys would not be useful as there are not many people who have a complete overview of all relevant (business) processes that ultimately lead to the DIS delivery, and who have insights in the organizational and technical problems associated with this. Surveys for example would not be useful for this reason.

3. Of qualitative approaches, the expert interviews are most useful. Two frequently used data gathering methods in qualitative research approaches are *participating observations* and *qualitative interviews* [13]. Clearly, observing medical specialists on their treatments (or following certain patients) and comparing this with how the data ultimately ends in the DIS, is outside the scope of this research, if not impossible. Furthermore, it is expected that there are different stakeholders with different views on the DIS data quality and possibilities. As experts can be included in a study such as this one not as a single case but as representing a group of people in a certain field (stakeholders) [34, p. 165], when sampling the right interviewees this will be a powerful tool. Interviewees would be able to provide answers to ”open” questions (semi-structured interviews based on quality dimensions), while representing a group of stakeholders.

2.2.1 Interview Protocol

The expert interviews are executed in a semi-structured way. An interview protocol has been written which is used to provide structure during the interviews. This interview protocol is based on Case Study Protocols as described by Yin [92] and is attached in Appendix A. Through this, it is ensured that the gathered information is comparable and not too much time is spent on information gathering not directly applicable to answering the research questions. The questions refer to three different aspects: intrinsic data quality approximated by identifying organizational issues and data quality in the context of ”fit for use” for efficiency and patient safety research. The interview protocol provides an outline for each of these aspects, describing the different topics that need to be addressed. Although in most cases all aspects were treated, most of the interviews were focused on only one of the three aspects.

Furthermore, the protocol provides a summary of the research (questions), although these differ slightly from the formulations in this final thesis' version. Some other data collection procedures are described including secondary sources, expected gathered evidence, how this will be used to answer the research questions, and validity threats. The high-level interviews outline (approach) to answer the different research questions are described below.

Interview Structure To assess intrinsic data quality, first the organizational processes that eventually deliver the DIS data (the data delivery process) are assessed from the interviewees. During this, some data quality issues are already identified. The ”standard delivery process” as described by DIS [20] is used to guide this assessment of hospital-specific delivery processes.

The remaining of the interviews related to the intrinsic data quality are performed semi-structured, by separately treating different intrinsic data quality dimensions. For each quality dimension, the interviewees provide the organizational issues they consider of influence on the final data quality dimension.

After each dimension is discussed with the expert and these issues are identified, he or she provides a grade between 1 and 10 (10 is highest) for this dimension. This grade is based on the overall opinion of the expert concerning the dimension and relates to a combination of the severity of the problems and the importance of the dimension to the overall data quality.

To assess the contextual quality, a comparable interview structure is followed both to the usefulness of the DIS data for efficiency and for patient safety research. First, the currently used (research) methods and datasets are identified. Then, the properties of the DIS data are examined in collaboration with the interviewees. With this, the possibilities and limitations of the data with regard to currently used methods and potential other methods are identified. Finally, five different contextual data quality dimensions are treated. For each of these dimension, the interviewees provided a short description of their opinion on the quality of the DIS data related to this dimension. Similar to the intrinsic data quality assessment, the interviewees also provided grading for the different contextual dimensions.

2.2.2 Expert Sampling

The interview protocol also describes some generic information on the different possible stakeholders concerned with the DIS data, from which experts can be selected. The three aspects of this research (data quality, efficiency, patient safety) differ in field of interest or business. Experts with substantive knowledge on one of the aspects, such as patient safety, are likely not to have much experience on other aspects such as efficiency research, even though this should ideally not be the case. Furthermore different groups of stakeholders concerned with the DIS data exist. These are for example health insurers, employers from DIS, and heads of hospital administration. One of the goals in the sampling was to include at least two experts from each stakeholder groups to allow for *triangulation* (see Section 2.4). Furthermore, as many stakeholder groups as applicable were included. The interview protocol (Appendix A) provides more detailed information on the different stakeholder groups and why some (such as the health insurers) are completely excluded from this research. The potential interviewees are determined using *purposive sampling*: participants representing specific characteristics are purposively selected [13] while ensuring the aforementioned criteria.

Flick [34] describes two different aims of sampling: width and depth. The first strategy ”seeks to represent the field in its diversity by using as many different cases as possible in order to be able to present evidence on the distribution of ways of seeing or experiencing certain things” [34, p. 123]. The second strategy ”seeks to further permeate the field and its structure by concentrating on single examples or certain sectors of the field” [34, p. 124]. Flick further describes that the different strategies should not be combined due to limited available resources (people, money, time, etc.) but seen as alternatives. Due to the explorative nature of this research, the ”width” strategy is more followed. This explains the reason for approaching as many different stakeholders as possible and comparing and combining the provided information.

Sampling Results Six different stakeholders have been chosen to interview: these are experts from the *Hospital Administration (HA)*, *Hospital Business Intelligence (BI)*, *DBC maintenance and DIS (DIS)*, *Hospital (Internal) Consultants (HC)*, *External Consultants (CON)*, and *Researchers and scientists (SCI)*. When processing the results, these codes will be used in conjunction with a number to pseudonymize the experts while containing information on their stakeholders’ position.

Four experts were specifically interviewed on more than one of the thesis topics: a total of eighteen experts were interviewed. Categorized on stakeholders' position, these were these: five experts from hospital administration; three experts from hospital business intelligence; three experts from DBC maintenance (2) and DIS (1); two internal hospital (safety) consultants; two external consultants; and three scientists. All but three of these experts have been working for over ten years in their field of knowledge. The three others had performed a specialized (PhD) research, or had such a function, that these were still very suitable for expert interviews.

The "Interview Results" section (5) provides more detailed information on the different interviewees and the specific stakeholders approached to answer the different sub questions. Information from that section that identifies the experts are not included in the final *public* thesis version for reasons of confidentiality.

2.2.3 Contacting the Experts

The experts, especially those to assess the topics of efficiency and patient safety, were found through a LinkedIn search (with a combination of the keywords "healthcare" and "efficiency" or "patient safety"). Based on the search results, the online résumés were compared after which potential experts were elected for each stakeholders group (at least two per group). The selected experts were proposed to the thesis supervisor from Deloitte who, in all cases, agreed upon the usefulness of the experts and in some cases could even already serve as an introduction to these experts. Some other potential interviewees, especially the heads of administration departments, were specifically introduced by this supervisor. One expert (patient safety researcher) was introduced by the Utrecht University first supervisor and three experts were assessed through my own network. These were the external consultants, a senior manager and a director of Deloitte, and two were hospital business intelligence (senior) project leaders. The latter introduced me to one internal hospital safety consultant.

background information for interviewees A document containing background information for the interviewees has been drafted. Although designed as a separate document, the content of this background information has been attached in Appendix B. The document contains a short summary of the research and the research questions. Furthermore, information on the DIS data structure (with the entity-relationshipmodel) and some background on data quality, efficiency and patient safety is included. The document is sent to all interviewees before the interview took place. Three experts did not receive the file due to the short notice of their interviews. During the interviews, the sent background document has shortened the introductory explanation and ensured consistency in definitions of the used concepts.

2.2.4 Structuring the phases of the expert contact

To structure the and visualize the progress of the interviews, a table has been constructed. The goal of this table was to not miss any (crucial) steps and have a quick overview on the progress of the results gathering phase of the thesis. Each row of the table contained each (potential) expert to interview, where each column contained a specific step on the processing of the interviews from the first contacting of the (potential) expert to the final inclusion of all results in the thesis. The exact columns of this interview progress table, thus, the steps that are performed for the data gathering phase, are provided

#	Interview Phase
1	First Contact
2	Reaction
3	Interview date planned
4	Sent Background information Document
5	Interview performed
6	Interview transcribed
7	Interview summarized
8	Added to LinkedIn
9	Sent thank-you mail
10	Sent summary of interview
11	Received and processed summary feedback
12	Included in thesis

Table 1: Interview phases used to structure and track the progress

in Table 1. Not all interviews have been summarized eventually as the complete transcripts did more suit the grounded theory approach, as explained in the next (sub) section.

2.3 Analysis of the results - Grounded Theory

The organizational issues influencing the (intrinsic) DIS data quality have not been discussed in scientific or business publications. Similarly, the quality of these data in the context (fit for use) of efficiency and patient safety research has not been examined yet. Therefore, as not much is known about ”the phenomenon” in the organizational and scientific literature, exploratory qualitative researches are likely to prove a fruitful strategy [29]. The semi-structured interviews with different field experts, as described in Section 2.2, will provide the data that will be used as input for the analysis.

The analysis will be performed with a method comparable to the *grounded theory* approach [73]. Grounded theory focuses on discovery of concepts and theory that have not been identified and developed earlier. Although this does not entirely fit the aims of this research, the grounded theory approach provides a systematic procedure to thoroughly analyze rich qualitative data. By following this approach, the aim is to identify as many (organizational) issues and possibilities of the DIS data from the interviews as possible. Furthermore, the grounded theory approach does not require predefined hypotheses but rather forms these based on the gathered data [24].

> *”Grounded theory methods consist of a set of inductive strategies for analyzing data. That means you start with individual cases, incidents or experiences and develop progressively more abstract conceptual categories to synthesize, to explain and to understand your data and to identify patterned relationships within it. You begin with an area to study. Then, you build your theoretical analysis on what you discover is relevant in the actual worlds that you study within this area.”* [24]

2.3.1 Interview Transcripts and coding the data

The first important phase in analyzing the gathered data in grounded theory methods is to code the data. This coding in short is the process of defining what the data are all about [24]. For this, parts of the original data are enriched with short phrases (codes) describing the contents of the part.

As expert interviews have been performed, these interviews need to be transcribed to enable the grounded theory approach and code interview fragments for later analysis.

Therefore, in case the interviewee agreed on this the interviews have been recorded (fourteen of the seventeen interviews). From these recordings, interview transcripts have been created, literally reproducing every word spoken by the expert in the interview. The literal transcripts have been written on twelve of the interviews. Two of the interviews proved less useful to the actual research goals and only relevant fragments of these interviews have been transcribed. This is not an unusual approach to be ”less overwhelmed with the interview results” [4, p. 37] and to keep the dataset relevant. The resulting dataset contained almost 200 pages of interview transcripts. More detailed information on the performed interviews and the resulting dataset is provided in Section 5.

Charmaz identifies two different approaches of coding: *line-by-line coding* and *focused coding* [24]. Line-by-line coding is more detailed and describes for each line of text what it is about. This helps to make decisions about what kinds of data a researchers needs to collect next. Focused coding can be seen as a later step in the process and ”refers to taking earlier codes that continually reappear in your initial coding and using those codes to sift through large amounts of data” [24]. Focused coding is less open-ended and more directed and selective than line-by-line coding. This latter approach has been adopted for this research as the research questions are already more directed, specifically towards organizational issues and DIS data (im)possibilities. Therefore, the line-by-line coding phase is not deemed necessary.

After transcribing the interview recordings, each transcript is thus enriched with (focused) codings and headings to describe the topic and content of specific fragments.

2.3.2 Categorization and Analysis

Towards answering the research questions, the next phase was to group the codes into similar concepts. With this categorization, certain codes are selected as having ”overriding significance in explicating events or processes in the data” [24].

For each question, all groupings (categories) of interview fragments relevant to the answering of this question were gathered. From this, a first composition of the general ideas of the interviewees on the topic has been drafted. Charmaz refers to this as *memo-writing* [24]. She also states that this directly leads to *theoretical sampling*, i.e., collecting more data to clarify the composed ideas and to plan how to fit them together.

For this research, the collection of more data based on earlier data collection and analysis was only necessary to a certain extent. After the initially planned fifteen interviews, ”only” two more were planned to fill information gaps in the analysis. Furthermore, after the seventeen interviews no new codings, categories and information were extracted from these new interviews, besides the ”information gaps” for which these were planned. This implies that *saturation* had been achieved and no new interviews needed to be performed: ”Researchers are allowed to stop with the data collection when no new information on topics relevant to the research is extracted from the analysis of new interviews” [13].

For the final analysis of the interview results (Sections 6, 7, 8 and 9), literal quotes from the interviewees have been used where possible. This ensures the reader has insight in the statements on which the final conclusions have been based. Furthermore, many quotes have proven illustrative for a certain topic or issue. As the interviews were performed in Dutch, Appendix C provides all English translations used in this research, sorted by interviewee, accompanied by their original Dutch phrase and a page reference to the (external) interview transcripts document.

2.4 Validity Issues

Yin [92] describes four different criteria for empirical research. These criteria are construct validity, internal and external validity and empirical reliability. This section describes for each of these four validity threats what they imply and the actions performed in this research aimed to satisfy these.

Construct Validity Construct validity refers to the extent in which the used concepts are operationalized and measured correctly. To achieve this, only well-established concepts should be used to construct the theories, or these should be defined sufficiently [92].

To secure this validity threat, no new concepts were introduced. The concepts that are used are adopted from (influential) literature sources and specifically defined, including a notion on variations of the definitions and why these variations are not used. The literature background section (4) is partitioned in the three main topics of this research: data quality, efficiency and patient safety. Each of these sub sections start with a thorough identification of different related concepts and their definitions. This aids in ensuring construct validity.

The measurement of the concepts, in the context of this research, mainly refers to the assessment of data quality. In literature, it is mentioned that both *objective* measurements (metrics; test-based) and *subjective* perceptions of the individuals involved with the data (process-based) should be used to assess data quality (e.g., [25, 42, 65]). This research only focuses on the subjective, process based, measurement. The reason for this is described above, in Section 2.2, and is mainly due to the inaccessibility of the DIS data and the lack of a reference set. To assure a correct data quality assessment, the data quality dimensions that were questioned in the interviews were adopted from literature. Furthermore, an interview protocol (including an interview outline) was constructed to ensure consistent gathering and using of data.

Internal Validity Internal validity refers to the possibility to establish causal relationships and distinguishing spurious relationships [92]. As far as this thread is applicable to this kind of research (a descriptive study based on expert interviews), it is minimized it by using data source *triangulation*. Different stakeholders are interviewed for each topic. Furthermore, the aim was to interview multiple (at least two) experts from each stakeholder group to enable comparison of the opinions of comparable stakeholders, which further secures the internal validity.

External Validity External validity refers to the generalizability of the results [92]. For this research, generalizability mainly refers to the extent in which the found DIS data quality reflects the quality of the data delivered by all Dutch hospitals instead only those directly examined. As described in this section, many different stakeholders have been selected for the interviews, also representing different

types of hospitals and organizations. Six university hospitals were visited (some multiple times), one top-clinical and one general hospital. Reason for this is the initial belief that university hospitals deliver more complicated care and the administrative problems that exist in general hospitals, also exist in university hospitals. Whether this holds is described in the discussion Section of this Thesis (Section 12.3). That section also further describes the generalizability of the research results based on these interviews is discussed.

Empirical Reliability Finally, empirical reliability refers to the ability to repeat the research, with the same results [92]. This section and the interview protocol describe the outline of the semi-structured interviews and the sampling decisions. With this, it should be possible to repeat this research. Considering different stakeholder groups and multiple interviewees per stakeholder group are approached, it is expected similar results are gathered as well. However, it should be noted that due to the qualitative nature of this research, the conclusions drawn (based on interviewees' perceptions) could differ.

3 The DBC System in the Netherlands and DIS

This section provides some background information in the Dutch DBC reimbursement system and the accompanied DIS dataset. First, some information will be provided on the origins and incentives of the system in Section 3.1. This will also compare the DBC system to diagnosis-related groups (DRGs), to provide a reference method to international readers. Furthermore, the successor of the DBC system (DOT) will be discussed. Section 3.2 provides more information on the DIS dataset and its contents.

3.1 DBC's: Diagnosis Treatment Combinations

The DBC system is the Dutch hospital funding model introduced in the Netherlands in 2005. DBC is a Dutch abbreviation of Diagnosis Treatment Combination (*Diagnose Behandeling Combinatie*). In this system, the hospitals are paid a case-mix based tariff (by health insurers) for the entire path (DBC) a patient goes through from the diagnosis of a problem, to the treatment of the problem, to the final discharge [40].

The DBC system changed the older functional funding model in which hospitals received a fixed budget for the various functions a hospital provided. In this older model, three budget components were distinguished which were parameterized to determine the total budget: an availability component, a capacity component and a production component. This model had been the standard for over fifteen years, but had some serious problems. The main drawback was that the hospital expenditures frequently exceeded the provided budget as the growing and changing demands for health care had resulted in the situation where the parameters of the model did no longer realistically estimate the real costs [40]. Furthermore, lack of countervailing power of the health insurers and the limited incentives for efficiency (hospitals got paid by expenditures anyway) increased the urgency of the health-care reform [71].

The DBC system does have more incentives for efficiency, as hospitals benefit by increasing efficiency within a treatment path since they get paid an established amount for the whole path in the new model. Additionally, health insurers can negotiate the heights of these prices of a percentage of DBCs: the products in the so-called B-segment. With these freely negotiable prices, managed competition was introduced which also stimulates increasing the health care efficiency [38]. The prices for the other DBCs (within the A-segment) are determined by DBC Maintenance (*DBC Onderhoud*): the organization that develops and maintains the DBC products. The percentage of the negotiable DBC prices (those in the B-segment) was 10% when the DBC system was initiated in 2005, grew to 20% in 2008 and 31% in 2009 [9]. The B-segment is still growing with a targeted amount to 65–70% this year (2011-2012).

The set of hospital products are in practice around 30.000 different DBCs, although they can theoretically be much more (around 100.000) [40]. These different DBCs were based on almost 700 product groups, developed using advanced data analysis techniques. These techniques were performed on a data warehouse consisting of approximately 1.5 million patient records from 27 hospitals, gathered over a period of 3 years [88].

3.1.1 Comparing DBCs to Diagnosis-Related Groups

The Dutch DBC case-mixed based hospital funding model is a variant on the Diagnosis-Related Groups (DRGs), although there are substantial differences. DRGs have been used in the USA for almost thirty

years and many other countries are currently also using a variant of this DRG classification system for hospital reimbursements. A comparison with DRGs will therefore contribute to the understanding of the DBC system, especially for international readers. An extensive comparison is provided by Hasaart [40]; this subsection is heavily based on her work. She finds three main categories to divide the differences of the two systems in:

Differences in starting point and intention The previous section described the starting point of the Dutch DBC system: the old functional budget system before the introduction of DBCs. This thesis will not further describe the older starting points (the older systems) of DRG using countries. Regarding the intention of the system, the main reason to introduce the DBC system in The Netherlands was to introduce (managed) market competition. This differs from the main reason for introducing the DRG system in (most) other countries, which was to control the rise of health care expenditures.

Differences in Structure and Procedures The DBC has a broader scope than the DRG system: it consists of a combination of diagnosis and treatment. Besides this inclusion of the treatment component, also outpatient clinical care is included. This results in far more possible DBC performance codes than there exist for DRGs. Furthermore, it is possible to have multiple (parallel) DBCs open for one patient, where the DRG system only allows for one DRG at the same time for the same patient [40].

One of the other main differences between the two systems is the moment of the DRG/DBC assignment. With the DRG system, a DRG is automatically assigned to a provided care path *after* the discharge, and manually checked by a medical coder; with the DBC system, the DBC is assigned at the moment of the diagnosis by the performing specialist (e.g., general practitioner). The DBCs are assigned mainly based on this diagnosis (and expected treatments); the DRGs are assigned mainly based on the cost, specifically the recorded nursing days.

differences in financial incentives Finally, Hasaart describes that the differences in the DBC structures and procedures, mainly on the points of parallel DBCs and more detailed coding, result in different financial incentives of the two systems. Overall, the risk of upcoding, choosing (or administrating) the most lucrative DBC or DRG regardless of the medical necessity, is higher in the Dutch DBC system [40].

3.1.2 DOT: DBC Towards Transparency

Seven years after the (official) introduction of the DBC system, a new system was introduced on the first of January, 2012: the DOT system. DOT is an abbreviation of *DBC Towards Transparency* (Dutch: ”DBC’s Op weg naar Transparantie”). Comparable to DBCs, it describes care paths, although now referred to as *DOT care products* (in Dutch: ”DOT zorgproducten”). DOT improves three important issues within the DBC system [40]:

From 30.000 DBCs to 4.400 DOT care products One of the major drawbacks of the DBC system was the huge amount (around 30.000) of different existing DBCs. This caused the system to be intransparent and fraud sensitive [38]. For the DOT system, this amount has been reduced from roughly 30.000 DBCs to 4.400 DOT care products [22].

Products are automatically deduced In the older system, the diagnoses and activities were registered a DBC was ”manually” deduced from this: an activity internal to the hospital. In the DOT system, this internal coupling is replaced by an automated deduction of DOT care products, by a web-based grouper external to the hospital. This grouper is an algorithm that determines the provided care product based on the registered activities and patient information. To do this, the algorithm uses a decision tree that applies to the provided healthcare (referred to as a DOT-tree, or *DOT-boom* in Dutch). This introduction of the grouper changed the path of registration to the final declaration, which is further explained in Section .

DOT products are ICD-10 based The DBC products that resulted in the older system, was separately developed in The Netherlands where each specialism developed their own classifications. Therefore, international comparison and data exchange was not possible and the specialisms had their own ”island” of DBC products which were in some cases very similar to each other. In the DOT system, the coding of the diagnoses will be based on the uniform ICD-10 coding system: the tenth edition of the International Statistical Classification of Diseases and Health Related Problems. The coding of the activities will also change to a more detailed coding system (from the currently used CTG-coding to the more detailed CBV-coding) [22].

These three main changes have some important consequences to the system. Although there are more possible diagnoses and activities (combinations) with the use of the ICD-10 (and CBV) systems, by using the grouper, less care products are possible. These products, however, are medically more recognizable. One of the reasons there were so many DBCs is that they were bound to certain specialisms, resulting in the observation that some further completely identical DBCs are stored as two separate DBCs only for a different specialism and there was no clear solution for multidisciplinary diseases. These products are now uniformed, regardless of the specialism (Dutch: ”*specialisme overstijgend*”), which greatly reduces the total possible care products [22]. The negotiations between care providers and health insurers are also expected to improve with these more uniform and specialism transcending products [40]. Furthermore, the products are homogeneous related to the costs- and workload: they are better cost predicting and include indicators to represent the complexity of the provided care (”*zorgzwaarte*”). Finally, because of the use of ICD-10, performance data can be compared on both local and international level.

Financial incentives The DOT system also influences the means of financing. During the time of the DBCs, hospitals still had a budget (in the A segment) to cover the expenses to a certain extent if the registrations were not in order. With DOT, all hospitals are aimed to be reimbursed on performance (Dutch: *”Prestatie bekostiging”*), based on all their delivered care products. This performance reimbursement particularly influences the university hospitals where the size of the (with budget protected) A-segment was relatively larger than in general hospitals. To prevent irresponsible risks in the transition to the new reimbursement system, the university hospitals work with a ”transition reimbursement model” for the years 2012 and 2013. This model does maintain the market dynamics (managed competition), but the hospitals receive compensation when confronted with too large negative differences with their turnover. With this transition model, the traces of the functional budgeting model that still exist in university hospitals, will completely disappear in 2014 [81, Section 4.4.1.]. Also for the ”regular” hospitals, that were already reimbursed based on the DBC system, the transition from DBC to DOT can have some serious

consequences. A recent Deloitte research already found that the market liquidity of the Dutch hospitals has decreased, as hospitals can only declare an activity after (on average) ninety days, opposed to sixty days in in the DBC system.

Where the DIS was used earlier to gather DBC data, now the DOT data will be stored in it. Regarding this research, there are likely to be some differences in the DIS data quality between the DBC and DOT system. The data delivery process has changed somewhat (e.g., with the addition of the grouper) which also influences some of the organizational issues affecting the data quality. This research will focus on the (potential) problems with the DOT system, and where applicable also describe the similarities and differences with the DBC system. This also implies that many of the problems will be theoretical, as DOT has only recently been introduced and not a lot of DOT data is available in DIS yet. However, these issues were in most cases also present with the DBC system.

3.2 The DIS Data

The DBC (and DOT) reimbursement system is (partially) activity-based. Therefore, the hospitals are required to have their administration in good order, especially on those activities relevant to this reimbursement. Based on these registered activities and the accompanying diagnosis, a care product is derived for which a prearranged reimbursement is received. In the DBC system, this derivation was internally performed and later validated; in the DOT system this is externally performed by the grouper.

The accompanying registration, containing all these diagnoses, activities and resulting care products, is uniformly defined and stored in the national DIS database [21]. DIS is an abbreviation of *DBC Information System* and refers both to the database itself, and the organization that governs this database. The DIS data has been used for the development of the DBC and DOT care products, and is currently used mainly for the maintenance of these products and the product structure.

3.2.1 Entities provided to DIS by hospitals

As the DIS database is nationally defined, an extensive description of the required delivery format of the *Minimal DataSet* (MDS) exists [20]. This document provides the logical data model (ERD) of the entities and their relations every hospital at least has to conform when delivering their data to DIS. Figure 12 shows this entity-relationship diagram. Some data fields are mandatory to deliver (which is validated on): these items are shown in a bold font.

The most important entities of the DIS data are the patient, his care traject(s) and the different sub trajectories performed in the traject(s). These sub trajectories are referred to as DBCs as well, similar to the entire system. The various activities performed in the subtraject are stored as well, as part of the ”delivered care profile”. Some more entities exist in the data model: all are explained below, based on the document describing the required delivery format [20]:

Patient (Dutch: *”Patiënt”*) For each treated patient, the hospital has to provide information to the DIS. This entity contains attributes such as name, residence, gender, date of birth and social security number (Dutch: *”Burger Service Nummer”*). This information will not (all) be directly stored in the DIS: Section 6 describes how this private information is first pseudonymized.

Care path (Dutch: *”Zorgtraject”*) A care path starts the moment a patient expresses a certain demand for care [20]. This care path can consist of multiple sub trajectories, which are described below.

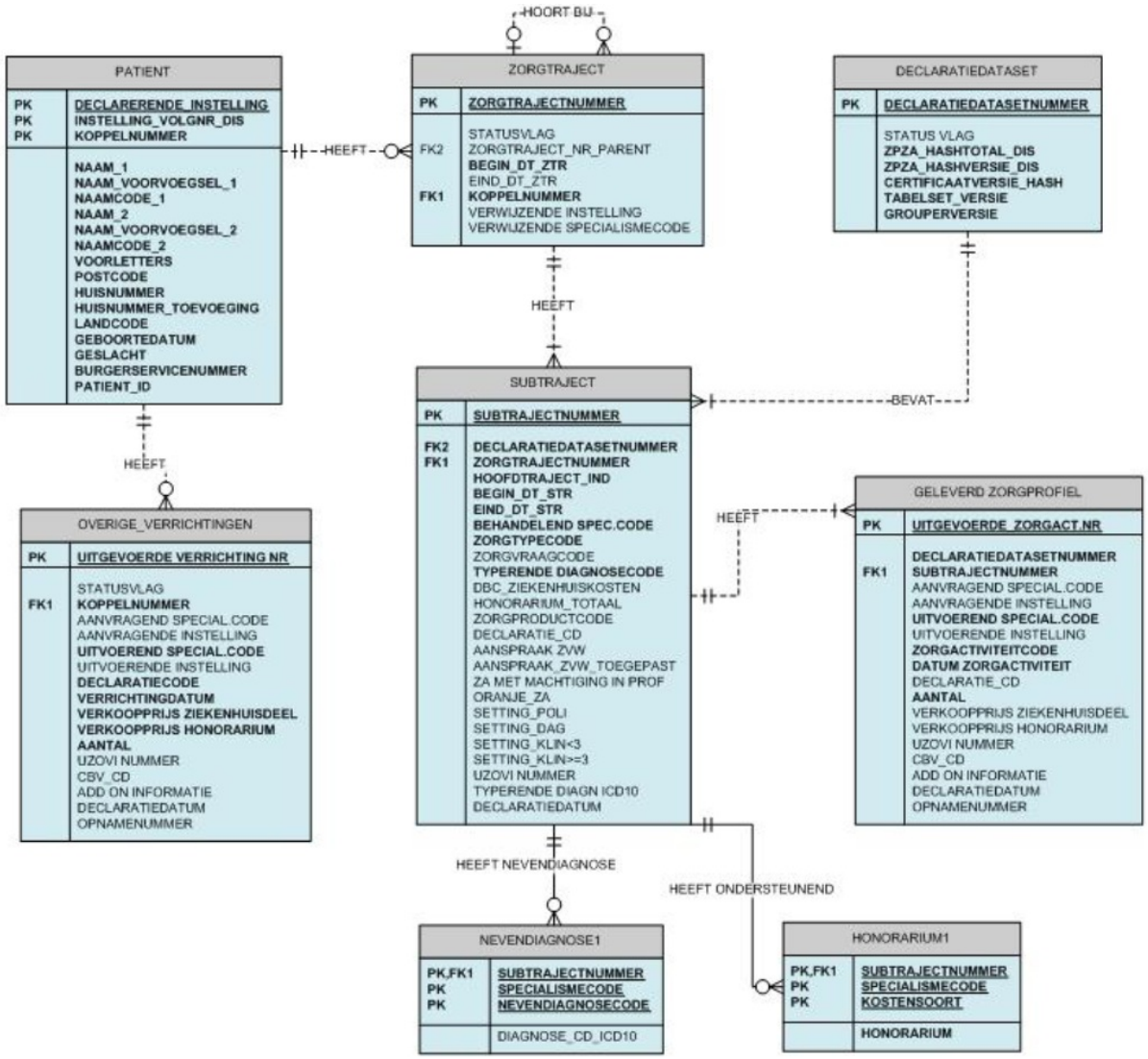

Figure 2: ERD Diagram of DBC Structure. Source: DIS (permission to use) [20]

The care path finishes when the care process (for a certain demand for care, or diagnosis) is finished. This also implies that some care paths (e.g., chronic diseases) could ”never” end. Some additional rules exist in these cases, which are not further discussed in this thesis.

Sub Trajectory (Dutch: *”Subtraject”*) The sub trajectories are the ”for declaration purposes defined part of a care path” [20]. This consists of a start and ending date and a unique classification for the type of care (e.g., initial or continued sub trajectory). Attached to it are the different activities performed in the sub trajectory which is described with the ”delivered profile” entity (see below). This sub trajectory is also referred to as DBC (diagnosis treatment combination) as it is the basic reimbursement attribute in the system: hospitals receive funding based on the performed sub trajectories.

Delivered Profile (Dutch: *”Geleverd Zorgprofiel* For each activity performed in a sub trajectory, a ”delivered profile” entity is constructed. This entity mainly includes an activity code to describe the activity and the number of times it is performed.

Declaration Dataset (Dutch: *"Declaratiedataset"*) The declaration dataset contains the information on the information that is sent to the DOT grouper for reimbursement purposes. This grouper creates a DOT care product based on the sub trajectory and its contents. This (DOT specific) information is stored in the declaration dataset entity.

secondary Diagnoses (Dutch: *"Nevendiagnose"*) Besides the (primary) diagnosis typical for one sub trajectory, it is possible that an amount (maximum seven) of secondary diagnoses were influencing the treatment. These can be attached with the secondary diagnoses entity, which is related to a sub trajectory.

Honorarium (Dutch: *"Honorarium"*) The honoraria for (certain) performed sub trajectories can be attached with the honorarium entity.

Additional Activities (Dutch: *"Overige Verrichtingen"*) Some care activities cannot yet be registered as part of one sub trajectory. These are usually reimbursed individually, and can be registered as "additional activities". These additional activities are coupled directly to a patient and thus not traceable to a care path or sub trajectory.

3.2.2 External (reference) tables

Many of the attributes in the DIS entities are unique identifiers referring to external reference tables (Dutch: *"Stamgegevens, Referentietabellen"*). This ensures standardization of the DIS information, minimizes free input and is used for validation of the delivered set by DIS. The DIS data delivery standard provides a complete overview of all reference tables [20, page 22-23]; the DIS website provides the full tables [21]. Information from these external tables can be used for analyses as well. Table 2 provides a description of the reference tables most relevant and important to this research.

Reference Table	Description
Provider of Care	*Instelling*: A unique identifier (AGB code) to identify the hospital that has performed a sub trajectory
Specialist	*Zorgverlenersspecificatie*: A unique identifier (BIG registration) to identify the specialist who has performed a certain care activity.
Health Insurer	*Zorgverzekeraar*: A unique identifier (UZOVI code) to identify the health insurer the patient is connected to.
Diagnoses	*El. Typeringslijst Diagnose*: The (primary) diagnosis attached to a sub trajectory is identified with a diagnosis code.
Activities	*Zorgactiviteit*: The activities performed by the specialist are translated into activity codes (ZA codes) and referred to as such. Section 6.1.1 describes this proces and its results in more detail.
DBC Product	*DBC Zorgproducten*: The final product for which a reimbursement from the insurer is received is typed with a DBC (DOT) care product code.
Rates	*Tarieven*: Certain reimbursement rates are nationally defined, for which a reference to the "rates table" is provided.

Table 2: Reference tables used to identify certain attributes in the DIS set.

4 Theoretical Background

This section provides an overview of the most important literature related to this research. Three main topics have been identified for this, related to the three sub questions of the research.

First of all, Section 4.1 describes data quality in general. Here, different data quality frameworks, measurement methods and improvement methods are treated. This will eventually lead to the composition of a framework applicable to this research. The framework will be used to guide the expert interviews towards an assessment of the intrinsic DIS data quality and organizational issues affecting this. Section 4.2 provides background information on efficiency. Related concepts are described, as are (healthcare) measurement methods and results of (Dutch) efficiency research. This will provide a basic knowledge needed to answer the second sub question on the usefulness of DIS data for efficiency research. Similarly, knowledge needed to answer the third sub question, on the usefulness of DIS data for patient safety research, is provided in Section 4.3. All concepts related to patient safety are defined and current measurement methods and results of (Dutch) patient safety research are described in that section.

The previous section (3) provided practical information on the Dutch DBC system and DIS. Combined with the scientific background information provided in this section, this forms the basis for this research and eventually answering the research question.

4.1 Data Quality

In recent years, the topic of data quality has gained increasing attention. It is recognized that *"data quality has serious consequences, of far-reaching significance, for the efficiency and effectiveness of organizations and businesses"* [7]. Furthermore, the reliability of the results when drawing conclusions based on certain data will decrease with bad data quality, e.g., *"deteriorate the process of discovering patterns, relationships and structures when applying data mining"* [7].

In many approaches to measure and improve data quality, no distinguish is made between data and information. Data quality and information quality are often even defined equally [41]. To ensure consistency with the (differences in the) definitions of data and information, and considering the dataset that will be examined in this research, this thesis will only use the term *data quality*.

4.1.1 Data Quality Dimensions

Many different studies have recognized that data quality is a multi-dimensional concept [65]. Many different data quality dimensions have been identified through literature, with many possible categorizations. However, no general agreement exists either on which set of dimensions defines the quality of data, or on the exact meaning of each dimension [6]. Therefore, choosing the different quality dimensions to measure the level of quality of data is the starting point of any data quality related activity [6, 7]. These data quality dimensions are not necessarily independent: some correlations can exist in them. A good example is that in order to improve the accuracy of the data, it might be necessary to build in multiple checks on this data, which would negatively influence the timeliness of the data [7].

Three data quality frameworks, each identifying different dimensions related to data quality, are described below. These are the hierarchical framework of data quality by Wang and Strong (1996) [85], the dimensions identified by Batini et al. (2009) [6], and the ISO 8000 standard for data quality [25].

The first (Wang and Strong) is important as it identifies the difference between intrinsic and contextual data quality, which will prove useful as this thesis aims to identify data quality in two different contexts (efficiency and patient safety). Furthermore, it is considered as one of the most influential researches on data quality (>1500 citations). The second (Batini et al.) is not really a framework but is important as it summarizes much data quality literature (including the Wang and Strong framework) and identifies the most frequently used quality dimensions. The last framework described, the ISO 8000, is still under construction. However, as ISO (the International Organization for Standardization) has published influential international management standards such as the ISO 9000 and ISO 14000, this data quality standardization effort is described here as well.

The Hierarchical Framework of Data Quality by Wang and Strong (1996) The data quality framework proposed by Wang and Strong [85] is one of the more influential (quoted over >1400 times) categorizations of quality dimensions. This framework, constructed in an empirical approach from 179 possible dimensions, categorizes 15 (originally 20) dimensions in four categories: intrinsic, contextual, representational, and accessibility data quality. The *Intrinsic* data quality describes the quality of the data itself, while the *Contextual* data quality considers the quality related to the context in which the data is used. Wang and Strong were the first to explicitly recognize that some of the data quality dimensions must be considered within the context of the task at hand [85]. The third category, *Representational* data quality, considers aspects related to the format and the meaning of the data. Finally, *Accessibility* data quality considers properties of data access, such as accessibility itself and access security [7, 85]. Each of these categories contain between two and five quality dimensions, which are shown in Figure 3 and further explained in the original research [85]. In a later research, the dimensions proposed in this framework have been evaluated in a hospital setting [36]. The conclusion is that these "quality dimensions are sufficient to define data quality in all sectors of the healthcare industry" [36].

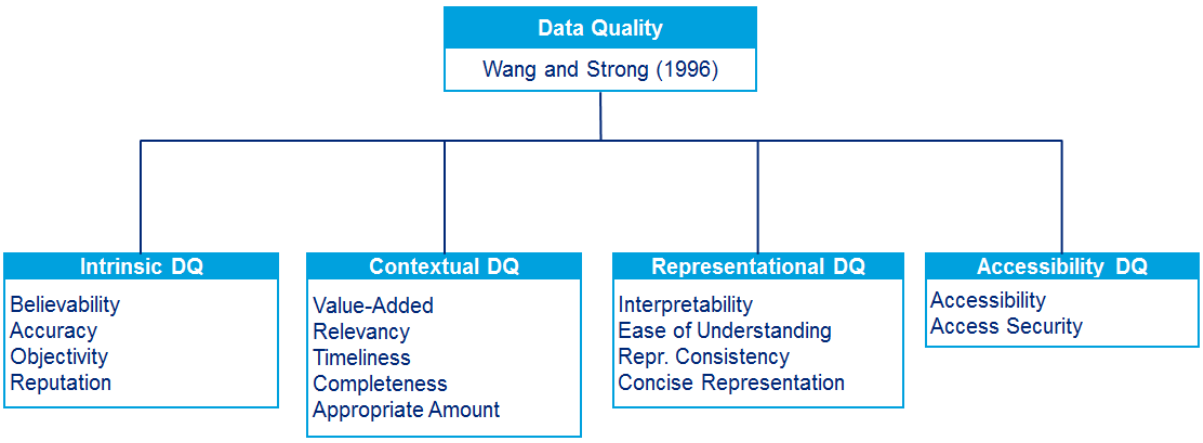

Figure 3: Hierarchical Framework of Data Quality proposed by Wang and Strong (1996) [85]

Lee et al. [57] summarize the different dimensions researchers have used to measure data quality, grouped in these four categories. This includes literature on data quality for specific fields, e.g. hard-copy reports and data warehouses. The research found two main differences when comparing these studies' quality dimensions. The first is whether or not *subjective* dimensions were included, based on whether the viewpoint of information consumers was considered. The other difference found

is that some quality dimensions could not easily be classified as being either an intrinsic or a contextual quality dimension. The example given in the paper is on the completeness and timeliness: *"as an intrinsic dimension, completeness is defined in terms of any missing value. As a contextual dimension, completeness is also defined in terms of missing <...> values used or needed by information consumers"* [57].

Dimensions identified by Batini and Scannapieca (2006) Batini has published two important works related to the quality dimensions: his book with Scannapieca in 2006 [7], in which an entire chapter is written on the different dimensions, and a publication (with Cappielo, Francalanci and Maurino) in 2009 [6] in which all influential data quality frameworks and (measurement and improvement) methods are compared. In the latter, much attention is also paid to the different definitions that are used in literature for the quality dimensions. Both works identify the same main data quality dimensions: accuracy, completeness, consistency and time-related dimensions.

However, in the book from 2006 also a subdivision of the dimension *completeness* is provided, based on the work of Pipino et al. (2002) [65]. Three different completeness dimensions are provided by Pipino: schema completeness, column completeness and population completeness. *Schema completeness* is defined as "the degree to which concepts and their properties are not missing from the schema", *Column completeness* is defined as "a measure of the missing values for a specific property or column in a table". Finally, *Population completeness* "evaluates missing values with respect to a reference population" [7]. This subdivision seems to solve Lee's observed problem, that mainly due to discrepancies in dimension definitions (Lee provided the example of completeness), some dimensions are difficult to place either in the intrinsic or contextual data quality category [57]. Schema completeness is typically an issue dependent on the context while column and population completeness are intrinsic to the data.

The final division of quality dimensions provided by Batini [6, 7] is shown in Figure 4. Besides these four main dimensions, all described in separate sections in his book, one section was on "Other Data Quality Dimensions". In that section, *accessibility* and *quality of information sources* were described, two dimensions that also exist in the framework by Wang and Strong.

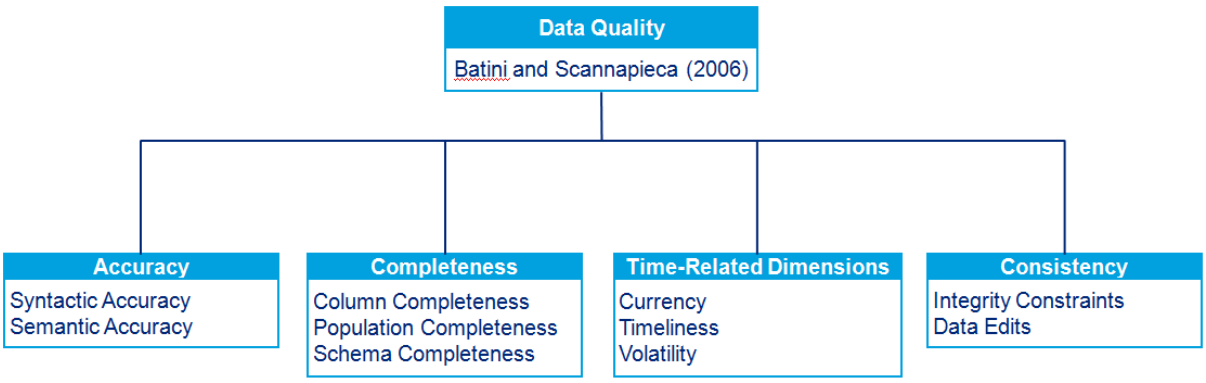

Figure 4: Data Quality dimensions identified by Batini and Scannapieca (2006) [7]

Where Wang and Strong provide four categories of quality dimensions (intrinsic, contextual, representational and assessibility), Batini and Scannapieca identify two different categories of dimensions. They describe that quality dimensions can refer either to the *extension* of data, which they also call

the data dimension, or to their *intension*, which they also call the schema dimension [7]. These two categories are comparable to the first two categories of Wang and Strong (intrinsic data quality and contextual data quality).

ISO-8000 As stated earlier in this section, no general agreement (or standard!) exists on the dimensions that define data quality and many different definitions are used for the same dimensions. Such standards do exist in other (quality assessment) fields. For example, for software quality the widely adopted ISO 9126 standard recognizes functionality, reliability, usability, efficiency, maintainability and portability as main software quality aspects. These aspects are further decomposed into sub-aspects. However, because of the different focus when assessing data or information quality compared to assessing quality of information systems, these aspects cannot simply be extended and used for this purpose.

Currently, work is being performed on creating such a standard: the ISO 8000 norm. This standard "*can be used by internal and external parties, including certification bodies, to assess the organization's ability to meet customer, regulatory and the organization's own data requirements*" [25]. The efforts on creating this standard started in 2006 and were initiated to improve the quality of data exchanged between organizations, in particular the quality of master data used to integrate the supply chain. This also results in the fact that the parts that are already finished are mainly focused in this area (especially master data). This also implies that not only a standard on data quality dimensions, but also on its measurement and improvement, will be created with this ISO 8000 norm.

4.1.2 Empirical Data Quality Assessment and Improvement Methods

Most data quality measurement methods found in literature are part of a larger methodologies that combine multiple activities. These activities are the assessment of data quality itself, identification of causes for bad quality, identification of root causes of discrepancies in data quality assessment results [65], and data quality improvement [57]. As mentioned earlier, this research focuses on the first two aspects, although there will also be some attention to improving the data quality by providing recommendations.

Up to 2002, mainly ad hoc techniques were available for measuring, analyzing and improving data quality in organizations [65]. With the recently increased attention to this field, more structured quality assessment methods have been developed such as AIMQ [57]. An excellent overview of these different measurement and improvement method is provided in 2009 by Batini, Cappielo, Francalanci and Maurino [6].

Many of the existing methods identify two different approaches of measuring data quality. The first is a comparative analysis of subjective assessments. This is approach assesses the subjective perceptions of the individuals involved with the data (e.g., through surveys and interviews). The second approach is performing objective measurements (metrics) [65]. The currently in development ISO 8000 norm also acknowledges this and describes two approaches to measure quality: test-based and process based [25]. The test based approach (quantitative) measures the product against certain specifications. This is relatively cheap to implement, but does not deliver process improvement. The process based approach (qualitative) identifies the problems in the business processes rather than in the data itself and tackles the quality systematically and delivers improved processes. However, it is stated that

this is (relatively) difficult, slow and expensive to implement [25]. Addico has developed and performed a process-centric approach applicable to the healthcare domain: this approach measures information quality in electronic health records [1]. Another process-centric healthcare data quality research is performed by Leitheiser in 2001 [59]. He has developed a process based model to measure data quality, specific to healthcare data warehouses, which allows practitioners and researchers to focus on processes that generate data quality problems. An example of a research that includes both quantitative (e.g., key value duplicates and referential integrity) and qualitative data quality checks (e.g., complaints from end users) is the one by Helfert, Zellner and Sousa in 2002 [42].

4.1.3 Conclusion and Date Quality Framework used in this research

With this background research on scientific data quality literature, two important aspects are concluded. First, multiple approaches to measure data quality exist and to this research. The next paragraph explains that a process-based approach is most suitable to this research and will be used during the interviews. Second, for data quality research, the first step is to create a quality framework describing the dimensions to classify the results and guide the research. This framework is created based on two of the frameworks presented in this section, which is further explained in the subsequent paragraph.

A Process-Based Data Quality Approach will be used for this research This section describes that data quality measurements can be quantitative (objective, test-based) or qualitative (subjective, process-based). Due to inaccessibility of the DIS data and lack of reference data, the quantitative approach is not possible in this research. However, data quality research does not only include measuring the quality, but also identifying causes for bad quality, causes for differences in quality measurements, and improving the quality. Although the quantitative approach provides a more fact-based and accurate quality estimation, the qualitative approach provides better understanding of the (root) causes for bad quality: this enables actual improvement of the data quality. The data quality measurement approach performed in this research will be process-based, both for this reason and because the test-based approach is not applicable. It will identify organizational (processes) issues that cause bad data quality, categorized with the data quality framework composed in the next paragraph. This will eventually aid in improving the data quality, for which recommendations are provided following the conclusions of this research.

Data Quality Framework composed for this research This section further described three data quality frameworks and that a framework applicable to a certain research has to be constructed on an ad-hoc basis. A framework is very useful in this research, due to the qualitative approach used. With it, the search for organizational (process) issues can be treated for each individual quality dimension, which scopes and guides the search for such issues. Therefore, a framework has been composed for this research. To ensure as many dimensions as possible are covered, it is based on an influential DQ framework (Wang & Strong [85], $>$1500 references) and supplemented with dimensions found in a book summarizing the dimensions mostly used in literature (Batini & Scannapieca [7]).

The Wang and Strong framework [85] is also tested in a healthcare setting [36] and identifies four different sub-dimensions of data quality (DQ): intrinsic DQ, contextual DQ, Representational DQ, and

Accessibility DQ. The focus (and scope) of this research is on the intrinsic and contextual data quality. The intrinsic data quality describes the quality of the data itself, while the contextual data quality considers the quality related to the context in which the data is used. In the framework, the intrinsic DQ consists of four sub-dimensions: Accuracy, Believability, Objectivity and Reputation. The contextual DQ consists of five sub-dimensions: Timeliness, Completeness, Value-added, Relevancy and Appropriate Amount. It is interesting to note that during the research of Wang and Strong, the completeness sub-dimension was first placed in the intrinsic dimension but moved to the contextual dimension after gathering and analyses of the results [85]. This seems consistent with an observation made by Lee [57], who states that some quality dimensions could not easily be classified as being either an intrinsic or a contextual quality dimension.

In the framework provided by Batini and Scannapieca [7], completeness is further divided in three sub(sub)dimensions: column, population and schema completeness (missing values related to a specific property or column in the table; missing values with respect to a reference population; missing concepts and properties from the entire schema). This seems to ”solve” this classification problem. Column and population completeness are typically dimensions intrinsic to the data, while schema completeness depends on the context the data are used for.

Similarly, the book by Batini and Scannapieca identifies three time-related dimensions instead of one *timeliness* dimension as identified by Wang and Strong. These time-related dimensions are currency, timeliness and volatility (how promptly are the data updated; the age of the data for the task at hand; the frequency with which the data vary in time). Here, currency and volatility are more intrinsic dimensions, while timeliness is more a contextual dimension.

Furthermore, Batini and Scannapieca divide accuracy in syntactic and semantic accuracy. Here, syntactic accuracy refers to the extent to which fields in the database represent *possible* real world states: for example, whether an activity code is an existing activity code. Semantic accuracy refers to the extent to which fields in the database represent the *actual* real world states: for example, whether the provided activity code describes the actual performed activity. Thus, Syntactic accuracy only determines whether a value X is in the domain D of possible X-values, semantic accuracy determines whether this value X represents the true value X' [7].

The framework constructed for this research is shown in Figure 5. It is based on the two (relevant) pillars of the Wang and Strong framework: intrinsic and contextual data quality. The dimensions within these pillars are supplemented with the sub-divisions provided by Batini and Scannapieca. Thus, syntactic and semantic accuracy are recognized instead only ”accuracy”. Furthermore, currency, timeliness (as in, age) and volatility are replaced for the original ”timeliness” dimension (from Wang). Finally, column, population and schema completeness are recognized instead only ”completeness”. As described above, currency, volatility, column completeness and population completeness are shifted to the intrinsic data quality pillar, resulting in the framework used to guide this research.

A final adoption to the framework was that the three intrinsic dimensions in the original Wang & Strong framework other than accuracy were grouped: believability, objectivity and reputation were aggregated to ”trustability”. This was after a strong relation and unclear overlap of these three dimensions was observed during the first interviews. Thus, this change to the framework was only made after these interviews.

The interview protocol (Appendix A) provides more information on how this final framework shown in Figure 5 was used to guide the expert interviews, and also provides definitions of the different dimensions as adopted from Wang & Strong and Batini & Scannapieca.

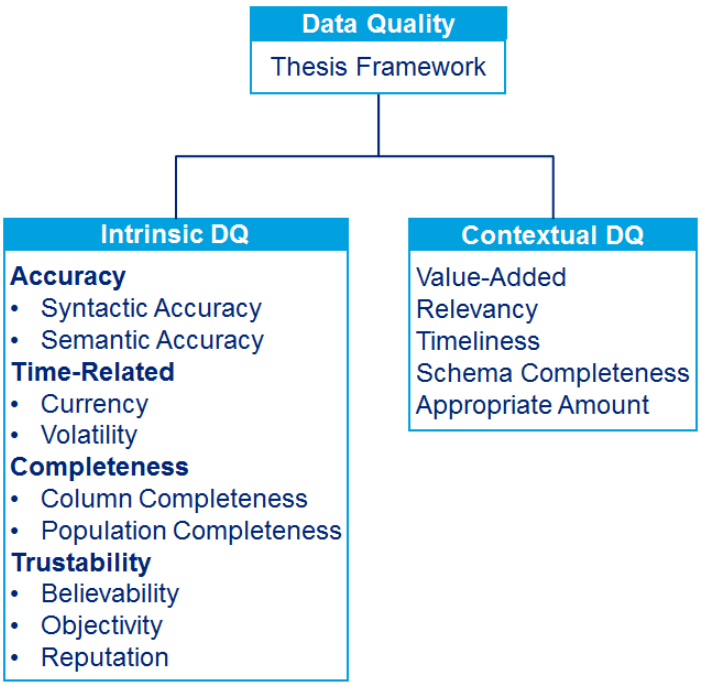

Figure 5: Data Quality Framework used for this research

4.2 Efficiency in Health Care

4.2.1 Main Concepts Related to Efficiency

In general, the term efficiency refers to the best use of resources in production [45], i.e., the relation between aspired performance and resources utilized [26]. Three main aspects of efficiency exist, which have been introduced almost fifty years ago by Michael Farrell and are still frequently used, whether in an adapted form or not. These measures are *technical efficiency*, *allocative efficiency*, and *economic efficiency* [32]. This methodology is also often applied to relate health care efficiency [45, 91].

Technical Efficiency Technical Efficiency refers to maximizing the amount of output from a provided amount of input, or similarly, minimizing the amount of input needed to produce a provided amount of output [45]. This also implies that it can not reduce all input factors proportionally, without reducing the output [60]. Within the context of health care, it refers to the *physical* relation between the resources used (e.g., capital, labor and equipment) and some health outcome [91]. Note that this brings up the question how to define the inputs and outputs of hospital organizations, a problem that is identified further in Section 4.2.3.

Allocative Efficiency Where Technical Efficiency refers to the ratio between the *total of all input factors* and the output, Allocative Efficiency refers to the *used mix of the different input factors* to produce an output. This Allocative Efficiency reflects the ”ability of an organization to use these inputs in optimal proportions, given their respective prices and the available production technology” [91]. In health care, when an allocative efficient organization (e.g., hospital) changes the *proportion* of the used ”inputs” (e.g. capital, labor and equipment), it will not improve its ”output” (some health outcome or financial measure), neither will this further minimize the total costs of the input factors.

Economic Efficiency Economic efficiency is defined as the product of technical efficiency and allocative efficiency [60], thus it is achieved when an organization uses its resources both technically and allocatively (completely) efficient. This implies that an optimal (cost-minimizing) mix of the different input factors is used and that this input mix results in a maximal possible output. Te term Economic Efficiency is also often referred to as *Full Economic Efficiency*, *Productive Efficiency* [91] or *Overall Efficiency* [45].

These concepts and their relation are explained further through a simplified case in which there are two different input factors X_1 and X_2 which result in one output factor Y. This case, which is a frequently used textbook example, has been adopted from the article written by Hollingsworth et al. [45] and is visualized in Figure 6. The horizontal and vertical axes represent, respectively, the X_1 and X_2 input factors.

In the Figure, the line TE shows all the technical efficient combinations of the inputs to produce one unit of output ($y = 1$). This line is also often referred to as the *Production Frontier* and a company operating on this isoquant would be *Technical Efficient* [45]. By definition, for each point (x_1, x_2) on this line there is no other combination of the two inputs with the same input sum, that deliver an output $y \geq 1$. The line AE shows all the optimal input-mixes of x_1 and x_2, thus that combination that minimizes the total costs while providing a same total sum of inputs. A company that operates on this line follows a cost-minimizing strategy.

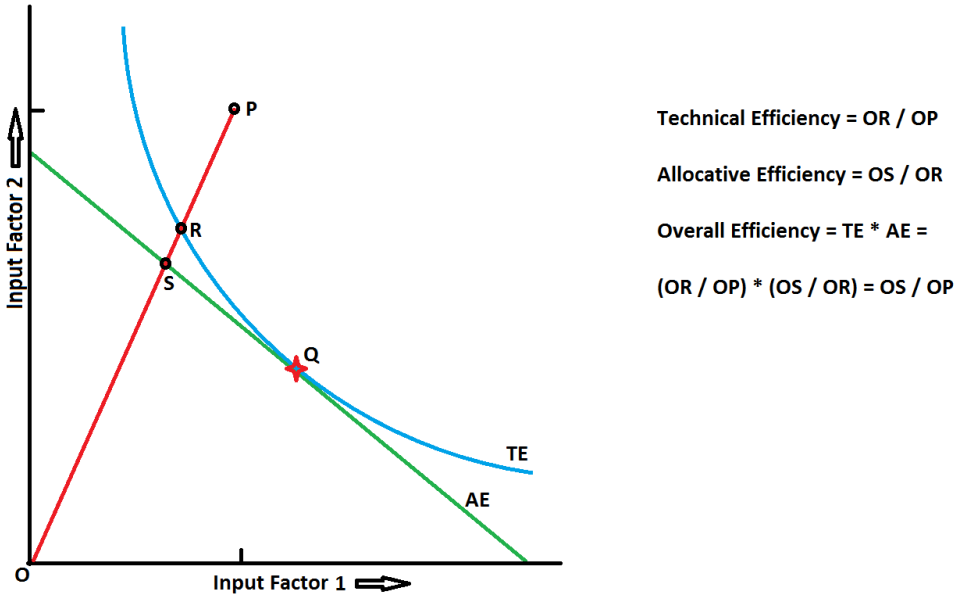

Figure 6: Relation between Technical (TE), Allocative (AE) and Economic (EE) Efficiency [45]

The combination of the two lines, point Q, shows where the output of $(y = 1)$ is achieved by minimizing the total amount of inputs, with the best possible (cost-minimizing) input-mix. A company that would operate on this point would be *(total) Economic Efficient*. This point Q for all different outputs (which would be a third dimension in Figure 6) forms the *cost or revenue frontier*, so a company operating on that on that frontier is *Economic Efficient*.

Now take a company that uses the input mix corresponding with point P, for which the output is also $y = 1$. The red line OP shows all other possible input mixes with the same allocation (proportion). Clearly, this company does not operate Technical Efficient: P is not on the line TE thus there are other productions with the same input mix, but lower total amount of inputs that result in the same output $y = 1$: point R would be optimal. Likewise, as the costs of the input mix used by P, would the total of inputs be technical efficient (point R), is not on the line AE (the cost-minimizing input mix), the company does not operate Allocative Efficient. There are other input mixes with the same total quantities of inputs as R, resulting in the same output (line TE), but that are more cost-minimizing: Q would be optimal. Furthermore, there are other productions with the same input mix as point P and R, but with yet other quantities of x_1 and x_2 (i.e., a different total sum of inputs) that are cost minimizing: point S.

$$TE = \frac{OR}{OP} \qquad AE = \frac{OS}{OR} \qquad EE = \left(\frac{OR}{OP}\right) \cdot \left(\frac{OS}{OR}\right) = \frac{OS}{OP} \tag{1}$$

To determine the level of efficiency of the company, the relation between the current situation and the most optimal situation has to be calculated. These formulas are shown at equation 1. For Technical Efficiency, this would be the relation between the lines OR (for the same mix, optimal input quantities to produce the same output) and OP (for the same mix, current input quantities to produce the same output). For Allocative Efficiency, this would only calculate the efficiency of the used *input mix*. Thus, it compares the lines OS (the lowest total costs possible with the used input mix) and OR (the total costs to produce a technical efficient $y = 1$ with the used input mix). The level of Economic Efficiency of

the company is defined as the product of the technical and allocative efficiency, which equals the ratio between OS and OP. Note that would a company operate at point Q, all the ratios from equation 1 result in (an efficiency of) 1.

4.2.2 Measuring Efficiency in Health Care

Many different methods to measure efficiency exist, from which two main approaches can be recognized: parametric and non-parametric approaches [91]. The differences in these approaches are based on a question arising from the definitions of the different aspects of efficiency: ”how can one determine the most optimal (cost-minimizing) combination of input factors (to calculate allocation efficiency) and the highest possible output accompanying a certain total input (to calculate technical efficiency)?” In other words, how are the cost and production frontiers constructed to measure the efficiency performance of the hospitals on? As this is usually not known, this most optimal situation (frontier) is constructed from the obtained data - which can be done either parametric or non-parametric. A third category of efficiency measurement techniques are non-frontier methods, although these have not often been applied in a health care context yet [45]. These three approaches are described below:

Parametric (Econometric) approaches Parametric approaches are also referred to as *Econometric approaches*. With this approach, the (efficiency) frontier is constructed so that it encloses all observed points (i.e., measurements of inputs with associated outputs). This frontier is constructed by fitting a parametric function to the data. This means that although all points lie within the frontier, they do *not* necessarily lie on the frontier and the *absolute* efficiency of the organizations are determined against some imposed benchmark (frontier). Some examples of this approach, which are also employed in health care research, are:

- SFA - Stochastic Frontier Analysis: When confronted with a deviation from the efficiency frontier, this method separates that deviation in two factors: inefficiency and randomness (or, statistical noise). This ”randomness” factor encompasses all aspects and events outside the control of the organization: an example is differences in operating environment. Furthermore, econometric errors such as misspecification of the frontier or measurement errors are part of this ”randomness” factor [91]. This method has been used by Ludwig to measure efficiency in Dutch hospitals [60].
- DFA - Deterministic Frontier Approach: Opposed to, and preceding the SFA method, this approach assumes *all* deviations from the constructed frontier represent inefficiency.

The main difference between the two methods described above is another dimension to distinguish the different efficiency measurement approaches on: stochastic versus deterministic approaches. Deterministic approaches assume deviations from the frontier as entirely due to inefficiency, where stochastic approaches also consider these deviations partly due to random error (white noise) [61].

Non-Parametric Approaches (Mathematical Programming) Non-Parametric approaches are also referred to as *Mathematical Programming*. With this approach, also an (efficiency) frontier is constructed so that it encloses all observed points. However, this frontier is not constructed through a parametric function ”around” the data, but by a ”non-parametric piecewise-linear convex frontier” [91]. This implies that the outcomes of some companies necessarily lie on this frontier and

form the ”efficiency standard” for the other companies, and the *relative* efficiency of an organization is determined against other, benchmark, organizations. Some examples of this approach, which are also employed in health care research, are:

- DEA - Data Envelopment Analysis: DEA is a method used frequently to measure efficiency (almost two thirds of 91 studies on efficiency before the end of 1997 made use of DEA alone [45]), and is also brought in practice to measure Dutch hospital efficiency [10]. It is a linear programming tool and calculates the efficiency of an organization in relation to comparable organizations, where deviations from the frontier are considered only to be the result of inefficiency [91].

  The technique is particularly appropriate when multiple inputs are used to produce multiple outputs, such as in health care [45]. However, as DEA uses information on input and output quantities, it can only address the issue of technical efficiency [53]. This has resulted in many variations of DEA (such as the less constrained *Free-Disposal Hull*) or combinations with other methods to examine other efficiency aspects [91].

- MI - Malmquist Productivity Indexes: When changes in productivity and efficiency over time have taken place, these differences can be measured with Malmquist Productivity Indexes [91]. An explanation of these indexes using a diagram similar to Figure 6 is described by Hollingsworth et al. [45].

DEA and SFA are by far the most frequently used efficiency measurement approaches in the context of health care, where DEA dominates the literature especially up to 1999 [45], but SFA is increasingly used since then [44]. The main differences of the two approaches are in the two dimensions of parametric (SFA) versus non-parametric (DEA) and stochastic (SFA) versus deterministic (DEA). However, a comparison of the approaches using the same data provided reassurance that both are measuring ”closely related constructs, along similar dimensions” [61]. The approach suited best to a situation is dependent on the design and goals of the study, a consideration every researcher has to make himself [60].

Non-frontier approaches (using average and bivariate indicators) Where the frontier approaches described above measure hospital efficiency performance against an industry best-practice, there also exist approaches that benchmark hospitals against average industry behavior. These approaches use bivariate indicators that take two summary dimensions (typically input/cost and output) into account. This also opposes the multivariate indicators of the frontier approaches, that evaluate the performance over multiple dimensions [61]. Two methods that are frequently described are *simple ratio efficiency* and *(least-squares or frontier) regression analysis*:

- Regression Analysis: Regression Analysis is a parametric approach, but rather than constructing a frontier (function) to determine the efficiency against, it constructs a function that best estimates the efficiency for all different observations (thus, a mean function). This means some observations have a higher efficiency (than, say, 100) and some a lower efficiency, opposed to efficiency values from 0 to 100 achieved with parametric frontier approaches (e.g., SFA). Some studies mention the construction of *frontier regression models* [5]; these are comparable to parametric (and stochastic) frontier approaches. The application of ”regular”

regression models in the context of health care are limited [45, 91] and not further discussed in this thesis.

- Simple Ratio Efficiency: Simple Ratios are performance measures based on simple cost/output ratios [61]. The emphasis of these performance indicators are often on inputs and processes, at the expense of outputs and outcomes, which is mainly due to the fact that the former are easier to measure than the latter [5]. Over Frontier measures, Simple Ratio Analysis has the principle advantage of simplicity. Although not all dimensions of interest might be captured, monitoring success of those dimensions that *are* captured is relatively easy [61].

Often, a combination of the different approaches described above is used. Examples of this are two stage analyses such as first performing DEA, followed by some form of regression analysis to further identify the determinants of efficiency after the measurement itself (mentioned in [45]), or consecutively performing DEA and business process simulation to analyze the efficiency of the business processes (i.e., throughput) [26].

4.2.3 Problems in Estimating Health Care Production Functions

Defining the output of a hospital is a notoriously difficult one [30], as improved health status or improved quality of life is difficult to measure [54]. Many studies therefore use measures of physical performance as output variables (e.g., patient days) instead changes in health status of individuals treated [45]. Another important reason for these output measurement problems is described as difficulties with aggregating dis-similar and ill-defined products, and the fact that multiple data elements of hospital care have to be taken into account to draw reliable conclusions [30]. Evans provides an example to the latter: with *only* data on days of care supplied, one cannot separate the days that were necessitated as a result of an adverse event from the days that resulted from poor organization of certain elements in the care process [30].

These problems with definition and measurement of the output also translates to input specification, although less complex than for specifying the outputs. Furthermore, the problem is eminent when determining the frontier isoquant or benchmark group and to a lesser extent (in The Netherlands) to the environmental constraints [60, 77].

4.2.4 Health Care Efficiency Research in The Netherlands

There have been many researches on efficiency in health care. These studies are mainly performed in the United States and several reviews provide a convenient overview of these researches and their results [44, 45, 91]. This section aims to provide an overview of the researches that have been performed in The Netherlands and also describes, where possible, used conventions in these researches to measure inputs and outputs.

Regarding this measurement of the inputs and outputs related to health care, two dissertations provide a solid base for Dutch hospitals [60], although these were written in 1977 and 1981 respectively and should thus be used with caution. First, Van Aert explains Dutch hospital cost functions [2] and second, Van Montfort [77] studied the applicability of production functions. The studies discuss the problems arising when estimating the production functions (see also Section 4.2.3) and point to several factors to measure input and output of (hospital) health care.

An early Dutch publication that really calculates the efficiency of the Dutch health care dates back to 1994 and was written by Peter Kooreman [54]. Although this research does not look into hospitals but into nursing homes, it provides a nice application of two different (non-parametric) DEA approaches: DEAC and DEACD. The first constructs the production (technical efficient) frontier while assuming constant returns to scale. This implies the frontier is a straight line, where an increase in input(s) results in a proportional same increase in output(s). The latter, DEACD assumes the production frontier to exhibit constant or decreasing returns to scale. This implies more institutions lay on the frontier, if they have increased input(s) in relation to the most efficient institution, but not necessarily proportionally increased output(s). A third approach, DEAV, expects variable returns to scale and forms the frontier by *all* best input-output ratios, but Kooremans study did not discover (significant) differences with the DEACD approach and thus omitted the results.

For all 320 Dutch nursing homes (at that time), data was gathered through surveys, from which 292 were useful. For these nursing homes, the technical efficiency with respect to the use of labor inputs was assessed using both DEAC and DEACD (and DEAV). Input variables were based on the numbers of (six different types of) employees, while output variables were based on the number of patients that were treated – many researches using the DEA approach use this to measure output. These treated patients were classified in four groups: full or day-care, for either physical or psychological disabled patients. The average DEAC efficiency of these homes was 0,80; the average DEACD efficiency was 0,94 [54]. One interesting thing to note is that Kooreman uses labor to measure output, while Montfort [77] uses labor as one of the three categories to measure input, amongst Capital and ”Other goods and services”. Besides Kooreman, also Jos Blank published on efficiency in nursing homes (in 1996) [10], where he found an average efficiency score of 0.70 [45].

In 1998, the cost and production functions of Dutch general and university hospitals were estimated by Jos Blank in the publication *Between Bed and Budget* [11]. In his later publication *Between Desk and Bed* in 2002, he described the measurement of efficiency in more detail, after which he published in 2004 on the cost structure and efficiency in Dutch general hospitals over the period 1985-1995 [12]. In this publication, estimations of both direct and indirect cost functions are made and found to be both plausible, thus, a mixed directindirect cost model was recommended. Both studies are based on an older Dutch national database constructed from annual national hospital inquiries on financial, personnel and production data (Dutch: *Enquête Jaarcijfers*).

Recently (2011), Blank published on the Productivity and Efficiency of Dutch hospitals between 2003 and 2009 (still based on the full version of the *Enquête Jaarcijfers*) [9]. It was found that the cost efficiency of the hospitals stayed constant around 0.92 and 0.93 through the years. The productivity was found to have increased with 14.7%. Furthermore, this paper includes some interesting findings: first, in order for Dutch hospitals to grow 1% in production, an increase of the costs of 1.23% (average) is necessary. This implies that when estimating the production frontier, decreasing returns to scale have to be assumed (as is the case with the DEACD variant of the DEA approach). Second, the research stated that the costs of a hospital could be explained well using only a limited amount of (product) indicators, while the DBC system increases these indicators significantly. It should be noted that from this econometric perspective (combined with the increased administrative burden), Blank is an opponent of the DBC/DOT system, which he also expresses in some publications (e.g., [8]).

Another recent study on Dutch general (non university) hospital efficiency is performed by Ludwig in

2009 [60], a researcher and practitioner who has also aided in the construction of the DBC system [88]. The study also uses the *Enquête Jaarcijfers*, with observations between 1998 and 2002. As input parameters the costs of staff and materials are used, which is not uncommon (a third frequently used parameter is capital costs, but Ludwig describes how this is exogenous in the Dutch health care sector and thus excluded from the study). Beside specific outputs for each department, five output parameters are used in this study to measure the output of hospitals. The first two are the number of trained specialists to measure the education (research is excluded) and the presence of an emergency department. To measure the health care provided, the other three parameters are used. These are the number of admissions, day-care admissions and first outpatient visits, where the first two are cost-weighted with a DRG (like) grouper. This grouper uses diagnosis and procedure codes and is, to a certain extent, comparable to the current Dutch DBC system. Ludwig uses the SFA approach and with it, finds an average efficiency of Dutch general hospitals to be 84%.

The information required in the dataset used in the Dutch efficiency studies (*Enquête Jaarcijfers*), has been reduced in 2010 [27]. This is a result of the DBC introduction and new national datasets such as the DIS. Although this decreases the workload of the hospitals, this also decreased the usability of this dataset for these kind of researches as some essential information is now missing. For this reason and because of the timeliness of hospital data (examined later in this thesis), the latest research on Dutch hospital efficiency is about data up to the year 2009.

4.2.5 Conclusion

This section has treated background information on the context of efficiency research, especially focused on hospital efficiency in the Netherlands. The first important observation is that three different gradations of efficiency measurement exist. These are *technical efficiency*, which relates to the rate of output that is provided from a certain amount of input; *allocative efficiency*, which relates to the rate in which the proportions of the used inputs (to produce an output) are in optimal proportions; and *economic efficiency*, which combines technical and allocative efficiency: the rate in which a company produces a maximal output from a cost-minimizing input mix. For determining the contextual (efficiency) quality of the DIS data, the gradation to which efficiency measurements are possible with the DIS data will be identified, based on the three aforementioned.

Furthermore, different approaches to measure (the different gradations of) efficiency have been identified. These differ from parametric versus non-parametric approaches; deterministic versus stochastic approaches; average versus frontier approaches; and approaches using multivariate versus bivariate indicators (e.g., input/output ratio versus more complex cost functions). In hospital healthcare, mainly frontier approaches are used: measuring efficiency versus a ”best-of-class”. The most commonly used methods are SFA (a parametric, stochastic approach) and DEA (a non-parametric, deterministic approach): both of these approaches are multivariate and frontier approaches. Other methods occasionally used in a health care settings are DFA (SFA but deterministic); MI (DEA measuring efficiency over time); Regression Analysis (like SFA parametric and stochastic, but measuring efficiency against an average instead frontier); and simple ratio efficiency (simple bivariate cost/output ratios). Which method mostly suits the needs of the specific research at hand depends on several aspects; the available data is an important one of these. When determining the contextual (efficiency) quality of the DIS data, the fit for use for the different measurement methods will be taken into account as well.

A last important observation made, often described in literature, is that quantifying outputs is an important issue when measuring hospital efficiency. Reason for this is that improved health status or quality of life is difficult to measure. Often, measurements of physical performance such as patient days are used. Specifying inputs (or, constructing cost functions) is also not straightforward considering the many possible input factors (such as used personnel), although less complex than specifying outputs as these input factors are by definition more quantitative. The DBC data mainly contains information on production (for reimbursement purposes). The possibilities for different input and output measurements based on these data will be taken into account when determining the contextual (efficiency) quality of the data as well.

4.3 Patient Safety

In 1999, the American Institute of Medicine (IoM), which advises that nation to improve health, issued the patient report *"To Err Is Human"* [51]. This influential report (over 8000 citations in twelve years) changed the view on health care quality by focusing on patient safety. The report started the discussion and focus on preventing health care incidents. Death counts as high as 44.000 to 98.000 were reported, directly due to (preventable) medical errors. These counts were higher than deaths by motor vehicle accidents, breast cancer or AIDS, supporting the need for this focus on preventing incidents.

Patient Safety is ensured in *"a situation in which the patient will not suffer, or has only a slight risk of suffering any damage caused by health care professionals who are not acting in accordance with the professional standards or by failures in the care system"* [83]. The *To Err is Human* report defines patient safety itself in this context as "freedom from accidental injury" [51], which refers to preventable adverse events [56]. To ensure this patient safety "involves the establishment of operational systems and processes that minimize the likelihood of errors and maximize the likelihood of intercepting them when they occur" [51]. These definitions imply that patient safety is obtained when the outcomes of the patient admissions are in line of expectations, thus where no incidents (especially those resulting in injury) have occurred.

4.3.1 Main Concepts Related to Patient Safety

In the different definitions, different names are given to health care incidents: these do not all refer to the same concept (e.g., incidents and adverse events are not the same). One of the results of this is that percentages of incidents (of total admissions) vary between different patient safety related studies, even though the same dataset is used (observed by [83]). In order to prevent this and to provide the reader a thorough understanding of the main concepts used in this research, this section provides definitions and explanations of the different concepts as they are most commonly used in literature. After the following list of concept descriptions, a Venn diagram (Figure 7) is provided to visualize the relation between the concepts, accompanied by a practical example.

Incident, Event, Iatrogenic Artifact An incident is an unintended happening that has led, could have led or could still lead to injury to the patient [87]. These unintended happenings have occurred during or following a medical intervention [43]. This is also referred to as iatrogenic: "Induced in a patient by a physician's activity, manner, or therapy" [28], a term derived from the Ancient Greek words *iatros* (healer) and *genesis* (origin). Thus, an incident is the same as an *iatrogenic artifact* (a term used in e.g. [67]).

In Dutch (and Belgian) research, usually only the term *incidenten* (incidents) is used; in international research, this is also often referred to as *event* to emphasize that all *adverse* events are incidents (but not vice versa). For this research, the term incident has been chosen because of the Dutch focus and the negative association with the term, emphasizing the need of improvement.

An incident can also refer to larger scale calamities, such as a bacteria outbreak. This research focuses only on smaller scale (patient level) incidents. Often, the larger scale incidents can be traced back to an incident on patient-level (e.g. not identifying the rare multi-resistant bacteria present at a patient).

Adverse Event, Sentinel Event or Critical Incident An adverse event is an negative outcome resulting from a medical intervention that, because of its severity, leads to prolonged or strengthened treatment; temporary or permanent (physical or mental) impairment; or death [87, 93]. This injury is not due to the underlying condition of the patient [51] (but iatrogenic). Another, comparable, definition for an adverse event that an adverse event is an unintended event resulting in harm caused by health care [83]. These definitions imply that every incident (or event) is a potential adverse event. An important addition to the definitions is that the injury does not *directly* have to follow the medical intervention, but a causal relation has to be evident [43]. Furthermore, it should be noted that, although all adverse events result from medical management, not all are preventable (i.e., not all are attributable to errors) [51].

Adverse events are by some official instances referred to as *Sentinel Events* [75], to emphasize the need for immediate investigation and response. In some official (law) documents, an adverse event is referred to as critical incident [39].

Medical Error A medical error is defined as "a failure in the treatment process that leads to, or has the potential to lead to, harm to the patient" [33]. This can be extended to the actions performed in the treatment process, where a categorization is made between an *error of execution*: a correct action which does not proceed as intended, and an *error of planning*: an incorrect original intended action [51].

However, not only these actions of individual hospital staff can cause medical errors, but all health care management: these individual mistakes as well as the broader systems and care processes [93]. Recent beliefs since the influential report *To Err is Human* [51] are even that the individuals should not be held responsible for medical errors and that the error can always be traced back to core problems in the process and supporting systems [43]. The latter is acknowledged in the latest report of the Dutch inspectorate on the state of health care [50], which identified that the information exchange using the different information systems was still inadequate and reason for many errors.

Avoidable (or Preventable) Adverse Events The American Institute of Medicine defines Preventable Adverse Events as medical errors resulting in injury [51], thus the intersection of medical errors and adverse events (see also Figure 7). Many studies prefer the term avoidable adverse events for this concept. It should be noted that not all adverse events (injuries resulting from medical intervention) are the result of a *medical error* and are thus not called avoidable. Furthermore, not all medical errors result in (serious) injury, thus not all are avoidable adverse events. Some studies (e.g. [66]) do interpret *all* health care errors (injury or no injury) as preventable or avoidable *adverse* events while they should technically be interpreted as preventable incidents. To avoid confusion, for this research the most common definitions will be used, as described in this section.

Figure 7 provides a Venn diagram in which the relations between the concepts are visualized. The sizes of the three sets shown in the figure do not represent the relative sizes of the concepts. The Venn diagram shows how both medical errors and adverse events are also incidents (but not the other way around). Furthermore, there is an overlap between adverse events and medical errors, which are the avoidable adverse events.

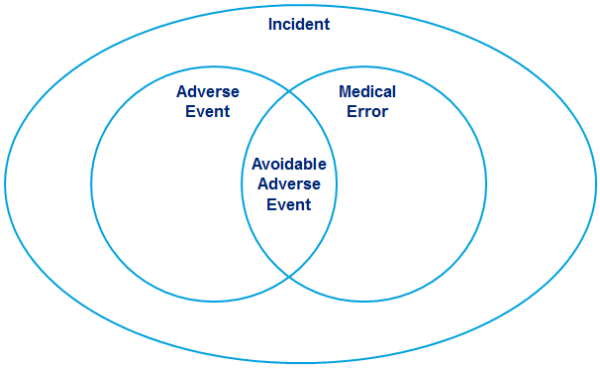

Figure 7: Venn Diagram showing the relationship between the different concepts (own creation)

To clarify the differences of the concepts with an example, fall incidents will be used. A study of three Midwestern hospitals derived an amount of 2,6 to 7,0 falls per 1.000 patient days [90]. All these falls are considered incidents, as they are all unintended happenings following the medical intervention that have the potential of inflicting injury to the patient. In not all of these fall incidents, the medical personnel can be blamed, thus these incidents would *not* be medical errors. An example would be when a patient, despite a provided warning, underestimates his local anesthesia in the legs or reduced vision and falls when attempting to go to the toilet. On the other hand, not all of the falls have severe consequences to the patient and are thus no adverse events. In the study mentioned earlier, these were 58-77% of the incidents [90]. In some cases of the fall incidents, medical personnel *can* be blamed: a medical error. For example, when a nurse leaves essential medicine out of a disabled patient's reach and the patient falls when going for his medicine. If such a fall incident resulting from a medical error results in injury (e.g., a broken leg), the incident was an avoidable adverse event.

There are some more concepts relevant to this part of the research that are not easily visualized in the Venn diagram. These are described below.

Injury or Adverse Outcome Injuries are consequences of the provided health services that are "serious and undesired, such as death, disability, injury or harm, unplanned admission to hospital or unusual extension of a hospital stay [39]. Some publications refer to injury as *unintended injury* [93], to emphasize the unexpectedness of the injury. Furthermore, injuries are sometimes referred to as *Adverse Outcomes* [46], although this could create confusion with the term adverse event. When looking at Figure 7, Injuries have occurred at all (preventable) adverse events; all other incidents " only" have the potential of injury. In Dutch literature, mostly the Dutch term *"schade"* is used, literally translated to *damage*. An important reason for this is that the term damage clearly includes more than only *physical* injury. Therefore, besides the term injury (often used in international literature), this thesis will also refer to this concept as *damage*.

Near Miss Near misses can be defined as incidents "(events) in which the unwanted consequences were prevented because there was an either planned or unplanned recovery by identification and correction of the failure" [46]. Incidents of which the consequences did not turn out to influence the physical, psychological or social functioning of the patient (although having the potential) are also called near misses, or sometimes *near incidents* [84]. When looking at Figure 7, Near misses are all incidents that did not turn out to be a (preventable) adverse event.

Complication A complication is an unintended result which has resulted in harm. This is a wider concept than adverse event as it does not only result from an incident, and thus it can only partly be found back in the Venn diagram (all adverse events are complications). Complications can also be the result of an unexpected reaction of the patient, such as an allergic reaction, or the result of a weighted risk or calculated risk [83] – thus are not necessarily a subset of ”incidents”. Calculated risks are defined by Langelaan et al. as a by the healthcare provider (e.g., specialist) factored (well weighted) risk or side-effect of which the gravity of the envisioned damage (injury) or the chance of its occurrence weights against the envisioned effect of the treatment [56].

4.3.2 Further classifications of Incidents

As defined, patient safety is closely related to (the absence of) incidents. The approach used for classification of these incidents commonly used in (Dutch) patient safety research is described above (and visualized in Figure 7). The starting point of this classification is binary: whether or not the incident was iatrogenic (caused by medical treatment), and whether or not the incident resulted in (severe) injury. Some further classifications than this one exist:

Differentiations in severeness of injury In many publications (e.g. [51, 90]), statistics of adverse events resulting in death or permanent (physical or mental) disability are separately provided. An adequate term definition for this concept is not provided in literature, but it is often referred to as *serious adverse event*; or it is simply described which serious injuries are included in the statistics.

Differentiations of iatrogenic actions Zegers [93] provides in his definition of causation three dimensions of injuries caused by health care management. He distinguishes *acts of omission (inactions)* (i.e. failure to diagnose or treat), *acts of commission or affirmative actions* (i.e. incorrect diagnosis or treatment), and *poor performance*.

Differentiations in type of incident In many researches, the type of incident is included in the results. For example, Ruben finds six categories to classify critical incidents based on former patient experiences [68]. These are, with the percentage of responses behind: Personal treatment/caregiver interpersonal communication (46,7%), Clinical/technical facets of the treatment (27,0%), Institutional policies and procedures (9,4%), Facilities and accommodations (7,3%), Quality/quantity of information provided (5,8%), and a category for other (3,8%). Another classification is provided by Gallagher and Kupas [35], who find ten main types of incidents, although very different from the one of Ruben. Another study, by Stella, Bartley and Jennings [72] differentiates two categories: management problems (69.4%) and system problems (27.1%). Besides the attention to organizational issues, also the latter, system problems, are identified in the Netherlands. The 2011 report on the state of the Dutch health care, written by the health care inspectorate, describes this. The report focuses mainly on the problems in communication between the different hospital information systems [50].

4.3.3 Measuring Patient Safety

As patient safety is defined as freedom from accidental injury, the first important step for its measurement is achieving incident information. More precisely, information on medical errors and (preventable) adverse events should be gathered. The first paragraph of this section is on how this incident registration is currently performed; the second paragraph is on current methods for analysis of these data, and alternatives when the data is not available.

Registration of Incidents One of the most important recommendations of the safety report *"To Err is Human"* firstly was the creation of an environment encouraging organizations to identify errors. For this, (mandatory) reporting systems would be installed. At that time, adverse event reporting systems were not or only partly mandatory in the United States [51] and other countries, such as the Netherlands [83]. After this recommendation also in the Netherlands a lot of attention has been on the design and implementation of quality (safety) information systems [15].

Information on mistakes made by hospitals and other unintended events, especially those resulting in injury, could harm a hospitals or hospital's employees reputation. Employees might be unwilling to share their observations of incidents, and hospitals unwilling to share this information. To prevent this, the "to err is human" report provided a second important recommendation. This was to shift the focus from blaming individuals for past errors to a focus on preventing future errors by designing safety into the system [51]. This should increase the willingness of employees to report incidents in the "new" reporting systems. Most incidents are often caused by a care process that has not been properly organized, even though at first sight they may seem to be caused by an individual [83]. Mistakes are easily made and could, and should, have been anticipated by adapting the process to minimize the chance for it. The study *Framework for analyzing risk and safety in clinical medicine* [79] supports this by providing its "anatomy of an incident", an accident sequence which traces back to the negative consequences of management decisions and organizational processes.

Analyzing Incidents An approach to analyze incidents is the PRISMA (Prevention and Recovery Information System for Monitoring and Analysis) method [70]. This approach also takes into account the previous mention: it does not aim to identify mistakes by individuals, but sees incidents as a result of failures within the organization (management). The method identifies the technical, human and organizational aspects behind an incident in an objective manner. Another approach used to analyze (already identified) incidents is performed in a study from 1998, where two thousand incidents were manually coded in a hierarchical, relational framework (GOC tree) [69]. A nine-step approach of analyzing clinical incidents, which also included structured staff interviews, is proposed and described by Vincent et al. [78].

In general, the analysis of incidents (and thus patient safety) is a powerful method of learning about health care organizations [78]. After identifying the errors through the mandatory reporting systems, evaluation of the errors would allow taking appropriate actions to improve future performances [51]. However, the incidents are not always directly available and gathering of incident information sometimes is the first step in a patient safety research.

An approach recently used in the Netherlands for this is performed by Langelaan et al. [56]. Here, complete patient records are analyzed. First, the likeliness of the occurrence of an incident is deter-

mined by nurses with the use of a trigger tool. This tool consists of sixteen indicators that might indicate the occurrence of an incident. The resulting ”high-risk” set is analyzed by medical specialists to identify incidents. Simultaneously, the severeness of the incidents (occurrence of injury and the severenss of the injury) is analyzed [56].

Automated incident discovery and analysis Besides these manual approaches, possibilities of automated approaches are currently explored. For this, patient records or related hospital data is directly and automatically analyzed for the occurrence of incidents and their severeness. In 1997, an article was published concluding that administrative data was at that time not suitable for quality (and related) measurements. However, the article was written fifteen years ago and already mentioned the fast developments on (availability of) these administrative data and described expectations many more possibilities in ”the future” [47].

For a specific category of health care incidents, adverse drug events, Jha et al. compared a computer-based monitor with the traditional approaches of expert (chart) reviews and voluntary reports [52]. One of the main findings of this report was that the computer-based monitoring system represented an efficient approach for measuring the frequencies of adverse drug events. In 2003, Hripcsak et al. published on the general possibilities of automated event detection and provided a framework for it. They concluded that this electronic incident detection appeared to be an important and feasible avenue of patient safety research [46]. In 2012, an Italian study (on Italian hospitals) evaluated a pilot surgical adverse event detection system. This system aimed to reduce the number of hospital records to be examined by medical specialists on adverse events, by removing records that certainly contained none. It was shown that, although continuous improvements to the system were required, the system indeed greatly improved the efficiency of the manual patient record reviews [16].

4.3.4 Problems related to Patient Safety research

Part of the reason for the current exploration of the possibilities of automated event discovery and analysis is related to disadvantages of other analysis techniques [46]. Although manual analysis of all cases by physicians is still used often for incident explanation [83], it is very expensive. For example, with the mentioned GOC tree approach coders could code eight incidents each hour [69]. This was already *”twice the rate as earlier techniques of assigning keywords”* [69]. However, significant time was still used as 2000 incidents and 800 adverse events were analyzed. Another drawback of manually (root cause) analysis of incidents, is that multiple interests and viewpoints of different stakeholders are involved, and the traditional methods are not good aligned with these [62].

Problems related to incident registrations Another important issue related to patient safety research refers to obtaining suitable and reliable data for the analysis, which was also an important issue with efficiency research. In this context, these are mainly problems related to incident registrations. Reports on incidents are crucial in many patient safety research approaches (some of which are described in the previous section); but this information is also for many internal hospital improvement programs required. However, caution should be made when interpreting results of (e.g.) error reporting as indicative of error rates [76]. This is especially the case before the ”To Err is Human” report, when not

much attention was paid to ”blame-free reporting” and focus was still on blaming individuals instead of organizational processes and management choices.

Before this ”To Err is Human” report, in 1991 only 5 to 15 % of medication errors were captured by error reporting systems [18]. Furthermore, earlier (see page 45) a paper written in 1998 by Jha et al. was mentioned [52]. An(other) interesting finding of this research was that only four percent of the identified adverse events were discovered through voluntary reports. Therefore, there was a need to be *”careful and recognizing that errors were frequently not documented in the medical record”* [46]. Although more recent international research has shown reports to be adequate for routine errors with immediate outcomes ([31], as cited in [80]), the safety management systems still do not provide information reliable enough to perform analyses on. An example from this is that in a study from 2005, the amounts of incidents reported in The Netherlands were very low compared to foreign studies. The main reason for this was stated as *”the voluntariness of incident reporting”* [83]. In 2012, a Dutch news site investigated that ”only” 243 (presumably) preventable deaths were reported to the Dutch Health Inspectorate [RTL News, April 28, 2012], where 1735 were expected ([56], see next section).

In an official (Dutch) report from 2006, it was recommended to keep incident reporting voluntary by law, but making it a moral duty. The most important aspect was to motivate care providers to report incidents and to create a safe environment for these reporters [58]. There are examples of (international) studies showing that the presence of mandatory reporting systems does not guarantee good incident reporting. One of these is a recent Swedish research of 100 cases has shown how adverse events causing severe harm were still to a great extent underreported in Sweden, despite the mandatory reporting system [63]. A reason noted was that physicians often considered these adverse events as complications and thus had no obligation to report them.

4.3.5 Patient Safety in The Netherlands

In the Netherlands also a lot of attention has been on patient safety, both in practice and scientifically. A report from 2004 describes how each year 1735 preventable deaths occur, a number frequently referred to in other literature and in the Dutch media. In that year, in 5,7% (76.000 patients) of the 1,3 million hospital admissions an adverse event took place [15]. 39,5% (30.000 patients; 2,3% of total admissions) of these incidents were (possibly) due to medical errors, thus preventable. Around 4000 patients died as a result of the adverse event (10,3% of a 3% death rate under hospital admissions). Between 1482 and 2032 of these (95% confidence interval) were preventable, thus the result of a medical error. The 1735 preventable deaths (mean) are derived from this: 4,1% of the total amount of deaths that year [15].

The study also calculates that the hospital stays are (on average) 9,1 days longer as a result of adverse events and even 10,3 days longer when these adverse events were preventable. These stays are more than double the average (regular) hospital stay in the Netherlands, which is (2004) 7,3 days. The direct extra costs of these extended stays were respectively 3844 euros (adverse events) and 4761 euros (preventable adverse events). The prices of the extra actions performed add on these amounts (around 1000 euros), as do prices of compensation to the patients (no data) [15].

A comparable study from 2008 shows that the percentage of adverse events has risen from 5,7% in 2004 to 8,0% in 2008. The percentage of preventable adverse events did not chance: although a grow from 2,3% in 2004 to 2,9% in 2008 was found, this was not statistically significant. The difference in preventable deaths, 4,1% in 2004 and 5,5% in 2008 (a grow from approximately 1735 to 1960), was not

significantly either [56].

Since 2004, hospitals are obliged to publish standardized performance indicators. These indicators were already used by a major Dutch newspaper to publish a ranking of almost all Dutch hospitals. In 2005, the Dutch government launched a health-care portal on the internet that could facilitate the health insurer choice for patients. This included information from the Customers Union [71] and was necessary as having a health insurer is mandatory in The Netherlands through the ”health insurance law”.

The Dutch report *Prevent damage, work safely* from 2007 [82] was the start of a safety program with the ambition to reduce preventable adverse events (and specifically deaths) with 50% in 5 years. One action to achieve this was the implementation of safety management systems (Dutch: *veiligheidsmanagementsysteem, VMS*) in hospitals and other health care providers.

Eighteen months later, the Dutch Health Inspectorate (IGZ) visited one quarter of all Dutch hospitals (considered a representative sample) and concluded that nice starts had been made with implementing these safety management systems [48]. All hospitals mentioned specific attention to patient safety in their policy and there was a system of safe incident reporting (Dutch: *”veilig incident melden”*).

In 2011, the Dutch Health Care Inspectorate (IGZ) published a list with the numbers of deaths in Dutch hospitals. These numbers were standardized against several factors such as the average age of the admitted patients and the type of care hospitals served. As a result, this list showed in which hospitals ”more patients died”. However, it was emphasized by the IGZ that this list could still not be used by individuals to assess the quality of a hospital, as this quality was caused by many other factors, and furthermore, the ”death numbers” were not standardized for all factors, although they were for many.

Besides research on the analysis of (quantities of) incidents in the Netherlands to describe patient safety, also researches have been performed on *prevention* of incidents. A good example is a recently published PhD dissertation that investigates some quality and safety improvement methods [55]. One of these methods is safety marking: markings indicating where certain tools or entire furniture have to be placed, which is adopted from the aviation industry to position planes. Some tools need to be placed in sterile airflow, which comes from fans at certain places in the operating room. Safety markings aid in correct positioning of these tools, thereby minimizing chances for incidents caused by unsterile operating tools [55].

Although such incident prevention interventions are also very interesting to analyze, this is not further discussed. Reason for this is that the focus of this research is on the identification of incidents *after* its occurrence and the possibilities of identifying causes of these incidents, instead of the process improvements to prevent future occurrences. With that being said, an ultimate goal of identifying incidents and their causes is obviously to improve the processes to minimize future occurrences, but this is not in the scope of this research.

4.3.6 Conclusion

The second context to determine the contextual quality of the DIS data in, is patient safety. This patient safety is obtained when the outcomes of the patient admissions are in line of expectations, thus where no incidents have occurred. This section describes background information on this context. An important observation made is that multiple gradations of incidents exist. The main concepts related to incidents and their relation are visualized in the Venn diagram on page 42 (Figure 7). As shown, incident is an ”umbrella concept”, containing all different gradations of incidents. There are two important

properties these incidents can contain. The first is whether or not the incident was iatrogenic (caused by medical treatment): in this case they are referred to as ”medical error”. The second is whether or not the incident resulted in (severe) injury: in this case they are referred to as ”adverse events”. Adverse events resulting from a medical error are referred to as ”preventable adverse events”. The main focus of current (Dutch) patient safety research is on the adverse events, and especially these preventable adverse events. Concerning the contextual DIS data quality for patient safety research, this thesis research is therefore also scoped to adverse events.

Furthermore, this section describes different approaches to measure patient safety, through the identification of incidents (adverse events) and analysis of the data. These are through analyzing incident registration systems output, ”manual” analysis of patient records and through automated event discovery based on administrative data.

For this research, more information on these different approaches will be gathered through the interviews. It will be examined to what extent (fit for use) the DIS data could serve as replacement for the different current methods. A focus in this aspect will be on automated incident discovery as the DBC (DIS) data is typically administrative data. Especially with the current issues related to patient safety research identified in this section, it is interesting to examine the contextual data quality (fit for use) of the DIS data to replace or complement the current approaches.

Part II

Results & Analysis

5 Interview Results (Public version)

Two versions of this results section exist: one for the final public thesis document, and one for the confidential thesis document aimed at the thesis supervisors. This version is the public version and contains far less detailed information on the experts, in order to ensure their confidentiality.

As described in section 3, expert interviews were performed to answer the research questions. In total, eighteen experts have been visited for this research (see Table 3). This was performed over a perios of 2 months. Some of the experts were interviewed on multiple of the research topics (intrinsic data quality and contextual data quality on efficiency or patient safety research); some were interviewed on one topic simultaneously in one group session. This has resulted in twenty-one interviews focused on one topic. Most interviews have been recorded and transcribed (or at least summarized), resulting in nearly 200 pages interview data. This dataset consists of the ”results” of this thesis research and is separately attached to the final thesis document. The analyses described in the following sections, and finally the conclusions and recommendations, are based on this dataset.

Nr	Function	Code	Main Interview Topic(s)		
Hospital Administration (HA)					
1	Head Administration	E1:HA	Data Quality		
2	Head Administration	E2:HA	Data Quality		
3	Head DBC / Medical Coding	E3:HA	Data Quality		
4	Projectleader DOT	E4:HA	Data Quality		
5	Head Administration	E5:HA	Data Quality	Efficiency	
Hospital Business Intelligence (BI)					
6	Projectleader BI	E6:BI	Data Quality		
7	Program manager BI	E7:BI	Data Quality	Efficiency	
8	Coördinator BI	E8:BI		Efficiency	
DBC maintenance and DIS (DIS)					
9	Manager DIS	E9:DIS	Data Quality		
10	Senior Analyst DIS	E10:DIS	Data Quality		
11	Medior Analyst DBC-O	E11:DIS	Data Quality		Pat. Safety
(Internal) Hospital Consultants (HC)					
12	Senior Consultant Safety	E12:HC			Pat. Safety
13	Quality Manager	E13:HC			Pat. Safety
External Consultants (CON)					
14	Director	E14:CON		Efficiency	Pat. Safety
15	Senior Consultant Healthcare	E15:CON	Data Quality		Pat. Safety
Scientists, researchers (SCI)					
16	Lead Efficiency Researcher	E16:SCI		Efficiency	
17	Efficiency Researcher	E17:SCI		Efficiency	
18	Safety Researcher	E18:SCI		Pat. Safety	

Table 3: Overview Interviewees and main interview topics

As stated, Table 3 provides an overview of the experts that are approached and the thesis topics (intrinsic data quality, contextual quality on efficiency, contextual quality on patient safety) that was focused on with the expert. The other topics were usually also treated, although briefly. The experts are grouped per type of stakeholder, as identified in the interview protocol (Appendix A) and in section 2.2.2. For example, all experts working in the field of hospital administration (e.g., the head of the administration department) are grouped accordingly. All experts also have a ”code” attached to refer to them in the remainder of the thesis. This code consists of two parts. First, the number of the expert (e.g., E2). Then, an abbreviation of his stakeholder group (e.g., HA for hospital administration). Table 3 further describes the function of the stakeholder. More information on the interviewees is not provided in this public version of the thesis for reasons of confidentiality. The confidential version (aimed at the thesis supervisors) also contains a section with more information on the function, background and relation with the topics of the experts that were approached.

Of the eight Dutch university hospitals, six have been visited. These were: (Dutch abbreviations) AMC, VU, UMCU, LUMC, Erasmus and UMCG. One basic hospital was visited (Bronovo) and one top clinical (eye hospital). The other locations visited were office locations: research institutes (Technical University of Delft and Nivel in Utrecht), Deloitte (Utrecht and Amstelveen) and the DIS / DBC Maintenance office itself (Utrecht).

More detailed descriptions of the interview results for each of the topics are provided in the following sections. Section 5.1 describes the interview results related to intrinsic data quality, Section 5.2 describes the interview results related to the contextual quality for efficiency research and Section 5.3 describes the interview results related to the contextual quality for patient safety research.

Nr	Interviewee	Duration of the Interview	Recorded?	Transcribed? Word count	Summarized? Word count	Interview Protocol followed?	Detailed Data path description?	Quality Dimensions Grading?
1	e1:ha	90m	✓	8.500	-	✓	-	✓
2	e2:ha	90m	✓	9.900	-	-	✓	-
3	e3:ha	90m	✓	8.300	-	✓	-	✓
4	e4:ha	75m	✓	7.700	-	✓	-	✓
5	e5:ha	90m	✓	7.900	-	✓	-	✓
6	e7:bi	75m	✓	*4.600*	2.100	✓	-	✓
7	e9,10,11:dis	90m	✓	7.800	-	✓	✓	✓
8	*e6:bi*	60m	-	-	800	-	-	-
9	*e15:con*	30m	-	-	400	-	-	-

Table 4: Descriptives of interviews related to Intrinsic Data Quality

5.1 Interviews related to the Intrinsic Data Quality

Table 4 describes the interviews that were specifically related to the assessment of the intrinsic data quality. The three interviewees from DBC maintenance and DIS that were interviewed in one session are combined (nr 7). Of the nine performed interviews, two should not be assigned the label of ”expert interview” due to lack of interview structure: the interviews were not recorded, did not follow the interview outline as provided in the protocol, and the quality dimensions were not graded. Seven interviews have proven specifically useful (by terms of input to the thesis): these were four heads of administration departments, one DOT project leader, one business intelligence project leader and the interview with DBC maintenance. One of these interviews did not specifically follow the protocol, although all topics were treated.

5.2 Interviews related to Efficiency

Table 5 describes the inverviews that were *specifically* related to the contextual data quality of the DIS data to measure efficiency (fit for use). Of the six performed interviews, three did not follow the protocol and no dimension grading was provided. This was due to the fact that these experts had more specific knowledge on the subject. Therefore, the protocol was not relevant as not all interview topics could be treated in those interviews. However, these interviews resulted in many insights interesting and useful to this research (especially the two with $>$1500 words summary). Furthermore, all the experts interviewed on intrinsic quality were confronted with the larger scope of this thesis, and provided much information on this topic as well.

Nr	Interviewee	Duration of the Interview	Recorded?	Transcribed? Wordcount	Summarized? Wordcount	Interview Protocol followed?	Quality Dimensions Grading?
1	e5:ha	45m	✓	4.500	-	✓	✓
2	e7:bi	60m	✓	*1.800*	1.100	✓	✓
3	e16:sci	60m	✓	7.000	-	✓	✓
4	*e8:bi*	45m	✓	*4.000*	1.900	-	-
5	*e14:con*	30m	-	-	800	-	-
6	*e17:sci*	75m	✓	-	1.600	-	-

Table 5: Descriptives of interviews related to Efficiency

Two of the efficiency interviews succeeded an intrinsic data quality interview, but the structures both to assess the data quality as the efficiency possibilities were followed as if the interviews were separate (these interviews took in total over 2,5 hours). The time and word count regarding the interviews in this table (5) only refer to the efficiency part.

5.3 Interviews related to Patient Safety

Table 6 describes the interviews that were specifically related to the usefulness of the DIS data to measure patient safety. Of the six performed interviews, two should technically not be assigned the label of ”expert interview”. These were shorter, more informal conversations of which more have been performed during this research (although not reported in this thesis). However, as some very interesting insights were acquired following these two conversations, a summary was made which was used for the analysis as well.

The four structured ”expert interviews” were with one information analyst of DBC maintenance, one safety and quality consultant (top clinical hospital), one safety manager (university hospital) and one researcher. The two other interviews were with external consultants (a Deloitte senior manager and a director).

Nr	Interviewee	Duration of the Interview	Recorded?	Transcribed? Wordcount	Summarized? Wordcount	Interview Protocol followed?	Quality Dimensions Grading?
1	Alexander Rengelink	60m	✓	7.100	-	✓	✓
2	Dirk de Korne	90m	✓	*4.000*	2.000	✓	✓
3	Max van Spronsen	90m	✓	6.300	3.000	✓	✓
4	Maaike Langelaan	90m	✓	9.700	-	✓	✓
5	*Machiel Westerdijk*	30m	-	-	400	-	-
6	*Bertien Dumas*	30m	-	-	300	-	-

Table 6: Descriptives of interviews related to Patient Safety

5.4 Resulting Dataset

The recordings of the interviews were used to create interview transcripts. This has resulted in a total of eleven complete transcripts and three transcripts only on relevant passages (the interviews with Erik van der Velde, Franck Asselman and Dirk de Korne). Four interviews were only summarized (one of the reasons not to label these as ”expert interviews”). The resulting dataset of these eleven transcripts, three partial transcripts and four summaries contained over one hundred thousand words on nearly two hundred pages: this document is separately enclosed to the thesis document.

In this thesis document, many quotes are used from these interviews (transcripts). Appendix C provides the original Dutch phrases which were translated for this research, and a page reference to the transcript bundle, to be able to read the context of the quote.

6 The DIS Data Delivery Process

As described in Section 4.1.3, this thesis research uses a process-based data quality measurement approach to understand organizational issues that negatively influence the quality, rather than providing quantitative quality measurements. This could eventually also aid in improving the data quality, rather than only measuring it. For this process-based approach, understanding of the relevant organizational processes is required. The close relation between these processes and the DIS data quality is also supported by a statement made by one of the interviewees:

> *"I will try to visualize the processes internal to the hospital and the delivery to DIS, then we will automatically get to aspects such as data quality"* [E2:HA][1]

This section describes these relevant processes, which involve all the steps from the generation (first registration) of the data internal to the hospitals, through how the data are submitted by the hospitals, to how it is finally stored in DIS (after pseudonymization and validation). These phases from registration to data storage are referred to as the *DIS data delivery process*. As mentioned, the DIS relates both to the national database which is subject to this research, as to the organization that gathers and validates the data for this database, and further distributes it. After the data storage (finishing the delivery process), DIS performs some further pseudonymization and flagging activities. The DIS is also responsible for further distribution of these data, concluding the life-cycle of the DIS data (as within the scope of this research).

A high-level overview of all the phases in this data life-cycle are visualized in Figure 8. This figure is created for the purposes of this thesis, based on information gathered during the expert interviews (especially those with [e2:ha], [e9:dis], [e10:dis] and [e11:dis]), the standard DIS delivery format [20, Page 9], and a description of the DOT processes from the DBC Maintenance organization [23, Page 5]. This figure also provides references to different subsections in this thesis, which explain the phases further.

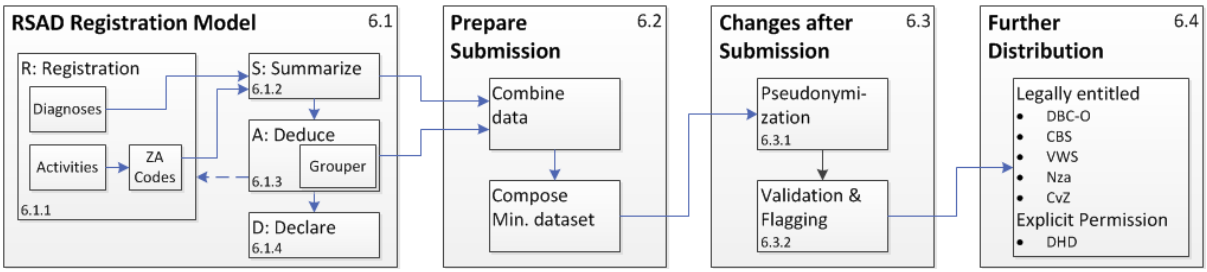

Figure 8: The DIS data delivery process and succeeding steps [20, 23]

[1] As this is the first time you read such reference, a short explanation: this abbreviation (E2:HA) is introduced in Section 5 and is used in the remainder of this thesis to refer to the expert from whom a quote was taken or based on whose statements a conclusion was drawn. In this case this is the second expert (E2). Furthermore, his or hers stakeholders' position is mentioned to relate the statement. Here, this is HA (hospital administration), as this interviewee is head of a Hospital Administration department.

6.1 The RSAD registration model

The first important phase of the DIS data delivery process is the sequence of activities to declare the provided care to the health insurers. As the DBC system has been developed for reimbursement purposes, these declarations are an important basis of the DIS dataset. Only the performed care for which a declaration has been sent to the health insurer, is eventually submitted to the DIS database [e3:ha].

The data processing involved in this declaration is referred to as the *RSAD registration model*. RSAD is the abbreviation for four words indicating the phases of processing: *Registration, Summarizing, Deducing* (Dutch: Afleiden)*, and Declaration* [23][e2:ha].

First, the data is *Registered* within the hospital: both the diagnoses and the treatments are performed and administered. After this registration, the activities are translated to care activity (ZA) codes needed for the later phases. This registration phase is further explained in Section 6.1.1. The next step is *Summarizing* the performed care, based on the registered activities and diagnoses. For this, the activities are combined with a leading diagnosis into a DBC (sub) trajectory, which is then closed based on standard closure rules. Section 6.1.2 describes this in more detail. The resulting information is used by an external, national grouper (algorithm) to *Deduce* a suitable care product (Section 6.1.3). The care product is returned to the hospital information system with some additional information, after which it is *Declared* to the health insurer based on earlier price arrangements. This concludes the RSAD process and is described in Section 6.1.4.

6.1.1 Step 1: Registration

The first phase of the RSAD registration model is the registration itself. Two different aspects related to the DBC (and DOT) system should be registered. Firstly, the performed activities, and secondly, the diagnoses. These are, at first, strictly separated: *"There is a bucket with diagnoses and multiple buckets with the activities: on the OR, policlinic, lab, pharmacy, I can go on for an hour. <...> But you see that the diagnosis, so the label on that DBC, and the activities, are separated from each other in buckets.* [e2:ha] It is important to note that, usually, all the hospital departments run a separate administration often even with separate information systems [e3:ha,e4:ha]. This explains the separate "buckets" of activities for those performed in separate divisions. The employees entering the data in the various systems (e.g., specialists, administrative employees, etcetera) are described in Section 7.8.1, which handles the trustability dimension of the intrinsic DIS data quality.

Besides the registration of the diagnoses and performed activities, the registration process firstly starts with opening a new care path, following the care request of the patient. With this, automatically a sub trajectory (See section 3.2) is opened as well [23]. Multiple sub trajectories can exist as part of a care path. These sub trajectories will be the basis for declaration.

Medical Aggregation to Care Activities For the deduction of care products by the grouper (the third phase of the RSAD process), the care activities should be referred to with "care activity" codes. Furthermore, this is the format in which the activities will eventually be stored in the DIS database. These "care activity" codes are identifiers referring to a Dutch national table with care activities (Dutch: *Zorg Activiteiten* or *ZA*), which are based on the older CTG codes used to identify activities [e11:dis]. In the remainder of this thesis, these "care activity codes" are referred to as *ZA codes*, to avoid confusion with "real" activities and other activity codings.

At this point, the registered activities are usually not stored as ZA codes yet [e2:ha]. Therefore, after the initial registration efforts, usually an aggregation has to be performed to translate the performed activities. For general hospitals (Dutch: *perifeer*), there are not many problems on this area, as these hospitals often still use the old CTG activity codes [e7:bi]. However, especially university hospitals register the performed activities internally with CBV codes. These CBV (Dutch: *Centraal Beheer Verrichtingenbestand*) codes are more detailed than the ZA codes. To illustrate: there are approximately 60.000 CBV codes which map to 4.000 ZA codes. This mapping can be one-to-one, many-to-one or no proper mapping could exist [e4:ha]:

1. **1 CBV code to 1 ZA code** Some ZA codes describe the activity performed by (e.g.) the specialist well. These map one on one.

2. **Multiple CBV codes to 1 ZA code** However, there are also more general ZA codes, on which multiple activities might map. For example, only one code for (placing a) pacemaker regardless of the type of pacemaker used exists. But the hospital might have pacemakers from different manufacturers in use with different properties (e.g., one with three and one with five wires). The costs of these pacemakers might differ, which would influence on the cost efficiency of applying the different pacemakers. On the other hand, there might be some higher risks with the cheaper pacemaker, which would influence the patient safety. But all these different pacemakers map to the same ZA code [e2:ha]. Another example is an MRI scan, which is one ZA code. But there exist more CBV codes that describe a MRI scan, e.g., one for the left leg, one for the right leg, etcetera [e4:ha].

3. **CBV code, but no ZA code (yet)** There exist situations where no (suitable) ZA code exists to represent a CBV code. Most commonly, these are codes that are used for research purposes which are usually reimbursed through other fundings: subsidies rather than insurers reimbursements (Dutch: *derde geldstroom*) [e7:bi]. Therefore, although university hospitals have an interest in the registration, these activities have no further influence nationally, or for DIS [e4:ha].

   However, also innovative activities exist, which are very common in university hospitals. Furthermore, there are many activities for specialized care. These categories of activities might also not have a suitable ZA code [e1:ha, e2:ha, e3:ha, e4:ha, e5:ha, e7:bi, e10:dis]. For the creation of these codes a dependence of the hospitals to DBC maintenance exists [e7:bi]. This situation occurs mainly in university hospitals and the problem is less eminent in general hospitals [e5:ha, e10:dis].

4. **New ZA code, but no internal CBV activity code (yet)** This one relates to the third problem. If DBC maintenance has created a new ZA code based on the hospitals' needs, the foundation responsible for the maintenance of the CBV codes has to create one such code so hospitals can register the activity. Furthermore, the hospitals (all the separate departments with the separate systems) need to be aware of the link between the CBV and ZA code, which in practice sometimes proves to be a challenge [e4:ha].

Besides these mapping issues, there is also the issue that some ZA codes are not directly mapped from one (CBV) activity but derivatives of multiple activities. An example is day-care (Dutch: *dagverpleging*) which is a treatment taking at least two hours, but not a full day (night stay) either. Both light and heavy day-care (ZA) activities exist: this is not registered (in the first step of the process) but a derived when the day-care activity is traced in combination with a certain class of operating room activities [e2:ha]. So this aggregation phase is both the translation of the care-logistics (CBV) code to ZA codes, and the execution of certain derivations. The influence of this aggregation phase in the delivery process on the final DIS data quality is expressed in the following quote, which was pronounced during the explanation of this aggregation step:

> *"You want to know something about the DIS. And you want to read things in those data. If you do not realize this <aggregations of registered activities to activity codes> you will think you read things in the DIS that are not there."* [e2:ha]

6.1.2 Step 2: Summarizing

The second phase of the RSAD registration model is summarizing the delivered care, based on the registrations. This phase consists of two activities. First, the diagnosis and activities are coupled to each other (i.e., the performed care activities (ZA codes) are connected to a sub trajectory with related diagnoses). Secondly, the sub trajectory is closed [e2:ha].

The mapping of all activities performed on a patient to a specific sub trajectory is performed based the execution date, type of activity, executing specialism and requesting specialism [23]. For this, a coupling algorithm exists in the hospital information system of the specific hospital. The closure of the trajectory is performed automatically following rules based on the activities in the trajectory. For example, 42 days after an OR (operating room) activity, the trajectory has to be closed [e2:ha].

6.1.3 Step 3: Deducing Care Products (the DOT grouper)

The third phase of the RSAD registration model is the deduction of a care product, based on the sub trajectory with attached diagnosis and treatment activities. This deduction is automatically performed by a web-based grouper, which operates external to the hospital.

This grouper is an algorithm which uses a decision tree that applies to the provided healthcare. These decision trees are referred to as a DOT-trees, or *DOT-bomen* in Dutch. The first "task" of the grouper is to determine the right decision tree: different trees exist for different groups of care products. This selection is made based on the diagnosis of the sub trajectory. After this, the grouper uses the care profile attached to the sub trajectory (the diagnosis and performed activities), where the (non)existence of certain diagnoses and activities determines which care product was delivered to the patient. [23].

This way, the grouper deduces the billable care products provided by the hospital. These care products (and possibly, certain add-ons to further weighten a care product) are then returned to the hospital. These billable products are also accompanied with a hashcode, which can be used by the health insurers and DIS to verify the grouper deduction of the resulting care products from the delivered care.

Rejected sub trajectories In some occasions, a registered sub trajectory is not deduced into a valid care product and is rejected. In Dutch, this is referred to as *"grouper uitval"*. These rejections cannot be declared and the hospitals have to review their delivery to the grouper and update it entirely. These rejections can occur in three situations [23]. The first situation is when the care profile is empty, or insufficiently filled. Here, the care profile refers to the performed activities. The second situation is when registration mistakes are discovered. These could be when activities completely irrelevant to the leading diagnosis are observed. The third situation is when no activities corresponding to the leading diagnosis are registered at all.

6.1.4 Step 4: Declaring delivered care

The fourth and final phase of the RSAD registration model is declaring the care products, which were deduced by the DOT grouper based on the summary of the registrations, to the health insurers. After receiving the billable care products from the grouper (together with a validation hash code), these products can be declared. For this, within the hospital information system the negotiated or set tariffs of the care products are combined to these products. Consequently, only products for which these tariffs have been negotiated can be declared. After this combination of the delivered care products and the tariffs, the products are declared at the insurer or patient for reimbursement.

When declaring a product at the insurer, the insurer does not require all information registered in the previous steps. Only information relevant to the declaration is in there: the performed activities are not in the system. This is also one of the reasons why no insurers were interviewed for this thesis: the DBC data they receive is far less detailed than the DIS data. The exact data(fields) required are described in the EI-standard[2]. The insurers store these declaration details in "their" national database: *Vektis*. As Vektis is the information bureau of health insurers, it is to a certain extent comparable to what the DIS database is for DBC maintenance, only less detailed [e5:ha].

6.1.5 The earlier DBC registration model: RVD

Section 3.1.2 describes how the DBC system is further developed to the recently introduced DOT system. The RSAD registration model described in this section relates to this newest DOT system. Before DOT was introduced in January 2012, another registration model was in place: the RVD model. This model was based on Registration-Validation-Declaration (R-V-D). In the registration and declaration phases, there are not many differences compared to the RSAD model.

However, the in DOT largely automated performed phases of Summarizing and Deducing, were in the RVD model mainly manually performed. After registration, the care paths and sub trajectories were closed manually, although limited to a maximal trajectory duration of one year. Furthermore, the billable care products were not automatically deduced by the grouper, but also manually determined. The validation phase mainly included validating these manual deductions. Often, a trajectory was closed "too soon" (the patient had to return with the same care demand). The "end-date" of the trajectory then had to be changed to be able to attach the newer activities to the profile and potentially deduce a "heavier" care product. Consequently, this frequently resulted in updates to the insurers and the DIS database.

[2] See "DBC/Ziekenhuiszorg" delivery formats on http://ei.vektis.nl/WespStandaardenOverzicht.aspx

6.2 Combining data and submission to DIS

Based on the summaries of the delivered care and the grouper output, the care provider composes the set to be delivered to DIS. This dataset is based on the *minimal dateset (MDS)*, which is described in Section 3.2.1 and visualized in Figure 12 (page 134). Remember, these data roughly include the performed care paths for each patient, where the care paths consists of multiple sub trajectories. The sub trajectories are the basis of the grouper deduced care products used for declaration. The delivered profile (the performed activities) for each sub trajectory is also included, as well as the deduced grouper information such as the declaration code of the care product and the validation hash code. This (minimal) dataset is submitted through an internal sending module, the PVM (Dutch: *Privacy Verzend Module*) [20].

6.3 Changes to the data after hospital submission

It is important to note that after submitting the (minimal) dataset, some modifications are made on the data before storage into the DIS database. Thus, the ER diagram shown in Figure 12 on page 134 is not the exact datastructure of the DIS database. Two important activities are performed on the data before storage in the database, each changing the data to a certain extent. The first activity is pseudonymization of the data, to ensure the privacy of the patients. This is further described in Section 6.3.1. The second activity is validation of the data and the addition of quality indicators (*flagging*). This activity is further explained in Section 6.3.2.

6.3.1 Pseudonymization

Based on the Dutch law for protection of detailed personal information (Dutch: *Wet op de Bescherming Persoonsgegevens: WBP*), not all patient information is allowed to be send to DIS and stored in the DIS database [20]. As shown in the database diagram of the minimal dataset (Figure 12 on page 134), the patient table contains a coupleNumber, *Burgerservicenummer* (comparable to social security number), full patient name, gender, postal code, address and country. The restricted part of these data concerns all information that can be used to deduce the identity of a patient. Therefore, the patient information in the hospitals' DBC data (the to-be-delivered minimal dataset) should be reduced in level of detail and some pseudonymization should be performed on "identifying" information. This data reduction and pseudonymization is performed in two phases, where a central role is fulfilled by a trusted third party (TTP): ZorgTTP.

1. For the first step of the pseudonymization, the data that cán be send to and stored in DIS are separated from the privacy sensitive information. The allowed data concern the couple number, which is used to couple the patient's data (table) to the other data (tables) and which is not related to the patient's identity information. Furthermore, the first four digits of the postal code, the country, year of birth and gender are not-restricted information [20]. The privacy sensitive information are the detailed name, full date of birth, full address and social security number. This information is stored in a separate file and pseudonymized. Some fields are combined for this pseudonymization step, which is explained in further detail in the DIS delivery format [20].

   For this data separation and pseudonymization, the privacy send module of ZorgTTP (*Privacy Verzend Module, PVM*) is used, which operates internal to the hospital as an extension of the hospital information system. The input of this module is the minimal dataset, the output is the dataset with reduced and pseudonymized patient data which is then encrypted and sent to the central privacy module: CMT (Dutch: *Centrale Module TTP*) for the second pseudonymization phase.

2. This second step of the pseudonymization is thus performed external to the hospital: the CMT operates at ZorgTTP itself. This central module decrypts the data and adds a second pseudonymization to the privacy sensitive information, which is DIS specific. Later on, this will be important as one of the organizations (the CBS, Central Bureau of Statistics) requires some of the privacy sensitive information: see Section 6.4. However, DIS is in no way able to deduce the personal data [e9:dis]. After the second pseudonymization, the two patient files (sensitive and public information) are combined, and together with the other DBC data send to the DIS.

Besides these two pseudonymization phases, a third ZorgTTP module exists: the DRM. This module runs at the DIS itself and ensures the receiving of the files. This module has some extra validation checks to ensure the files were indeed sent from a hospital(through the PVM and CMT modules). After receiving, this module ensures the extraction of the files after which the data is available to DIS.

6.3.2 Validations, database storage and flagging

After receiving the data through the three ZorgTTP modules (which ensures the pseudonymization and safe transaction), the DIS performs three steps surrounding the database storage. These are a first validation on the delivery format, a more thorough validation of the data format, and enriching the data with quality flags [e9:dis, e10:dis]. All the validation tests detailed in the DIS data delivery format [20].

- The first step is the validation of the conformance with the data delivery format: the correct files with the prescribed file namings and index numbers should have been delivered. If this is correct, the file containing the patient (table) information is looked into, to determine whether the (right) pseudonymization has been performed. Once this first data delivery format is validated, the data are stored in a temporary database (”staging-in”).

- After this initial database storage, more detailed validations on the data format are performed. These validations include uniqueness of primary keys; secondary key linkage; filling of required fields (not null, no empty strings); correct relation multiplicities; and correct data types (e.g., date fields should only have valid dates stored, integers should be integers). Furthermore, the linkings to external tables are validated (see Table 2 on page 24: e.g., an activity code should be an existing activity code in the DBC maintenance list of ZA codes) [e10:dis]. For the data registered with the older DBC registration model (the RVD model, see Section 6.1.5), the DIS rejects an entire dataset if less than 97% of the data are positively validated. Furthermore, the (max 3%) ”invalid data” is *not* stored in the DIS database, and always returned to the hospitals [e9:dis]. With the recently introduced DOT system, part of the data have been generated by the DOT grouper (the RSAD model, see Section 6.1). For these data, the DIS recognizes validation ”warnings” and more serious ”errors”. The whole dataset is rejected by DIS in case of one error. Reason for this is the added grouper functionality, which already ”catches” most errors. Those cases would already result in rejected care products, which can not be declared and which are thus not sent to DIS (see Section 6.1.3). If the data pass the DIS validation tests (for 97% or 100% depending on whether the grouper was used to deduce care products), it is stored in the ”real” DIS database.

- After the (definite) database storage, quality flags are added based on the contents of the data. This process is referred to as *flagging* and enriches the data. An example of such a flag field is described in the following interview quote:

  > *”we know certain DBCs are clinical treatments so we expect treatment days. Or one of those activities. If that is not the case, we put a flag there”* [e10:dis].

  These ”quality flags”, or, potential quality issues, are returned to the hospital. The hospital should then determine whether or not there is a real issue with the data, although this is not mandatory [e9:dis].

6.4 Further distribution of the DIS data

There are a couple of organizations who receive the DIS data for analyses. These organizations are either legally entitled to the data, or have explicit permission from all the hospitals they receive the data from (the individual hospitals still have ownership of the data). This section describes the most important receivers (users) of the DIS data [**?**].

DBC Maintenance, DBCO As mentioned earlier, DBC Maintenance (Dutch: *DBC-Onderhoud*) uses the DIS data for the further development and maintenance of the DBC system. Furthermore, health care providers, insurers and governmental organizations can request DBCO to perform changes in the DBC system. DBCO reviews these requests (internally and externally) and potentially implements these changes. A good example of such a change can be the development of a new care product (a new possible grouper deduction), which enables the declaration of an innovative treatment.

Central Bureau of Statistics, CBS The CBS uses the DIS data for statistical analysis of the Dutch healthcare industry. For these analyses, often more detailed information is required than what is stored in DIS after the pseudonymization (Section 6.3.1). For this reason, the two pseudonymization steps exist. The following quote from the interview explains how with this, (only) the CBS is able to use more detailed patient information:

> *"The second pseudonymization key is typical to DIS and always the same. The CBS also has pseudonymized data with a CBS pseudonym. If we deliver to the CBS, it goes through ZorgTTP. These remove the DIS pseudonym and place the CBS pseudonym. Through this for example the social security number (SSN), although still pseudonymized and thus not directly recognizable, can be coupled to the SSN pseudonym the CBS already has."* [e10:dis]

Other organizations legally entitled to the data Besides DBCO and the CBS, there are three other organizations that are legally entitled to the data for the execution of their legal tasks. These are the Dutch Ministry of Health, Welfare and Sport (VWS: *Ministerie van Volksgezondheid, Welzijn en Sport*), the National Healthcare Authority (NZa: *Nederlandse Zorgautoriteit*) and the organization for healthcare insurers (CvZ: *College voor Zorgverzekeringen*). The uses of these organizations of the data are further explained on the DIS website [21].

Dutch Hospital Data Besides the organizations that are legally entitled to the data, the DIS also provides data to third parties that have explicit permission to these data from the hospitals. Dutch Hospital Data (DHD) is such an organization. DHD gathers hospital information from different datasets (e.g., also from the LMR and EJ mentioned earlier in this thesis). DHD also serves as the main party to request DIS data.

The DIS also provides periodical overviews back to the hospitals. These are high-level production volumes of the specific hospital, based on the data that was eventually stored in the DIS database [21].

These different receiving organizations have different views on the DIS data quality, which is mainly caused by their different uses (contextual quality) of these data [e10:dis, page 39-40 of the interview transcripts]. Besides DBC Maintenance, these organizations do not use the data in the contexts of

efficiency and safety research, the contexts examined in this research. Furthermore, as focus of this research is on the process-side of the data quality, the expert interviews were mainly performed with stakeholders on the data generation side, and not with the parties described above (except DBC maintenance).

7 The Intrinsic Data Quality of DIS from a Process Perspective

As stated in Sections 2.4 and 4.1.3, a process-based approach will be used to evaluate the intrinsic quality of the DIS data. This approach aims to identify organizational issues which eventually influence (bad) data quality. The previous section described the DIS data delivery process, based on an analysis of the interview results and two other important sources ([20], [23]). This delivery process is the most relevant one to the eventual DIS data quality. With this process information, this section provides an analysis of the interview results on the intrinsic quality of the DIS data. This analysis will aid in answering the first sub-question: *"What is the intrinsic data quality of DIS and how is this determined by organizational and technical issues?"*.

During the interviews, to identify these issues and estimate the data quality, different quality dimensions are used. These are described in Section 4.1.3 and shown in Figure 5 on page 31. For approaching the intrinsic data quality, these were four quality dimensions with a total of nine sub-dimensions. First, there were syntactic and semantic accuracy. Then, two time-related dimensions: currency and volatility. A third time-related dimension, timeliness, was issued often as well although it is more a contextual quality dimension. Furthermore, there were column and population completeness. Finally, three dimensions related to the trustability of the data were treated: believability, objectivity and reputation.

Each sub-dimensions was raised separately during the interviews: this section is split accordingly in sub sections where each sub section describes one sub-dimension. An exception is for the "trustability" dimension: there was a lot of overlap (or unclarity of the boundaries) between the different sub-dimensions of trustability. Therefore, these are treated simultaneously in Section 7.8.

For each sub-dimension, the interviewees provided their quality perception of that specific dimension and the organizational issues related to it. These quality perceptions and issues, where possible based on quotes stated in the interviews, are summarized in this section. In some occasions, the experts also provided "their" solutions to the issues, which were also described here and used as an important sources for the recommendations. To conclude treating the quality dimension, the interviewees summarized their perception of the data quality and the severeness of the (bad) quality in one grade (1-10, 10 highest). The results of these gradings are shown in Table 7. This table provides a first impression on the expert opinions on the different dimensions.

Intrinsic Quality Dimension	e1:ha	e3:ha	e4:ha	e5:ha	e7:bi	Average	St. Dev
Syntactic Accuracy	9,0	9,5	5,5	9,5	9,0	8,5	1,7
Semantic Accuracy	6,0	4,5	6,0	7,0	5,5	5,8	0,9
Currency	4,0	6,5	8,0	5,5	8,5	6,5	1,8
(Timeliness)	-	5,5	-	7,0	8,5	7,0	1,5
Volatility	6,0	6,0	8,0	9,0	8,0	7,4	1,3
Column Completeness	8,0	6,5	6,0	9,0	8,0	7,5	1,2
Population Completeness	7,0	5,5	4,0	8,0	6,0	6,1	1,5
Believability	6,0	8,5	-	9,0	7,0	7,6	1,4
Objectivity	6,0	5,0	-	9,0	9,0	7,3	2,1
Reputation	6,0	8,0	-	9,0	8,5	7,9	1,3

Table 7: All gradings related to the quality dimensions

Differences in the DBC/DOT system Most of the organizational issues identified relate both to the original DBC system (with the RVD registration model) and the DOT system (with the RSAD registration model, with automated closure rules and added grouper functionality). The differences between these systems and registration models are explained in more detail in Sections 3.1.2 and 6.1.5. For the analysis of the intrinsic data quality, the DBC system is taken as starting point as there is not much DOT data delivered yet due to the recent introduction. Where the interviewees identified potential (different) issues influencing the data quality (differently) due to the DOT system, this is separately mentioned.

7.1 Accuracy: Syntactic Accuracy

The first sub-dimension of intrinsic quality assessed is syntactic accuracy. This refers to the extent to which fields in the database represent *possible* real world states: for example, whether an activity code is an existing activity code (whether it is the right activity code is a matter of semantic accuracy).

Regarding this syntactic accuracy of the data there are no real threats. As one interviewee said: *"The whole healthcare sector is combined with tables, you see that in the datamodel"* [e5:ha]. Most of the fields in the DIS database are standardized and based on external tables (as described in Section 3.2.2). There is hardly any free input and the DIS database only contains data that is validated on this prescribed input [e9:dis] (see also Section 6.3.2). The good syntactic accuracy also shows from the grades provided by the experts on this subject (Table 7): these are mostly positive ($\mu = 8.5$; $\sigma = 1.7$). One interviewee (e4:ha) scored this dimension much lower (significantly raising the standard deviation), which was due to confusion of definition: the expert was far more positive on this quality dimension after further explanation, although no new grade was acquired. Table 8 illustrates the overall expert opinion on this quality dimension with some relevant interview quotes.

Expert	Quotes Related to Syntactic Accuracy
e1:ha	*"The code in there describing the specialist, that specialist exists <...> there are no strange things in there"*
e3:ha	*"When it does not conform the syntaxes, it will not even get into <the DIS set>"*
e4:ha	*"To conclude, after validation not <no syntactic issues>"*
e5:ha	*"It is all validated and an activity code should be from a table. <...> There is almost no free text in there, it can hardly go wrong."*
e7:bi	*"When extracting the data some things might go wrong but these are rejected: <syntactically> it is of good quality"*
e9:dis	*"The invalid <data rows> are not stored"*
e10:dis	*"<there is validation> whether a code that is being used, an activity code: it should be an existing activity code. <...> Only when this is all OK, it is stored."*

Table 8: Interview Quotes concerning Syntactic Accuracy

With the DOT system (The RSAD registration method, see Section 6.1), there should be even less problems. Due to the grouper functionality, most of the syntactic issues are already identified when trying to deduce a (valid) care product [e2:ha, e10:dis] Furthermore, as described in Section 3.2.2, the acceptance rules of the data based on the DIS validations have been tightened with the DOT system (from 3% invalid data allowed to only warnings but no errors allowed) [e10:dis].

7.2 Accuracy: Semantic Accuracy

The second accuracy related sub-dimension of the intrinsic quality examined in this thesis research is Semantic Accuracy. This refers to the extent to which fields in the database represent the *actual* real world states: for example, whether an activity code describes the actual performed activity. There is a large difference between the semantic accuracy treated in this section and the syntactic accuracy treated in the previous sub-section. Syntactic accuracy only determines whether a value X is in the domain D of possible X-values, semantic accuracy determines whether this value X represents the true value X' [7].

Table 7 (page 65) describes the grades the experts have provided on their perceived quality of the DIS data of this dimension. Compared to the grades of the other dimensions, the semantic accuracy grades are lowest and most consensus exists ($\mu = 5.8$; $\sigma = 0.9$): the DIS data is considered to have most issues on this aspect. Some interview quotes further illustrating this, related to the overall perception of semantic accuracy of the experts, are provided in Table 9.

Expert	Quote Related to Semantic Accuracy
e5:ha	*"There are many more risks in that dimension <Semantic Accuracy>"*
e7:bi	*"The last two years, you can better assume the activity code represents the actual performed activity"*
e7:bi	*"Although semantic accuracy has increased, there are still question marks related to it. These points and examples will always be used by hospitals to undermine analyses on these data."*
e10:dis	*"These kind of mistakes, intrinsic to the DBC system, are not easy to identify and dissolve"*

Table 9: Interview Quotes concerning Semantic Accuracy

Two important issues were provided by the interviewees that explain the problems concerning the semantic accuracy of the DIS data. These issues occur in different phases of the delivery process. First, the codes used in the DIS database to represent diagnoses and especially activities are not detailed enough to represent the medical reality. Therefore, when aggregating the registered hospital activities to the national activity codes, which was already introduced in Section 6.1.1, often a lot of information is lost. This potentially negatively influences the semantic accuracy of the DIS data, which is explained in Section 7.2.1. The second issue relates to the deduction of a DBC/DOT product through summarizing of the care paths (coupling diagnoses and treatments) and the grouper functionality: these were already introduced in Sections 6.1.2 and 6.1.3 and are further described in Section 7.2.2.

Upcoding Besides these two issues, which are mainly the results of technical and organizational limitations due to the system itself, the issue of upcoding exists. Upcoding refers to deliberately changing a diagnosis or adding activities to receive (possibly undeserved) "heavier" care products and higher reimbursements [40]. This would influence the semantic accuracy of the data as it does no longer properly represent the real world state. Thus, it could be an expected organizational issue related to the semantic accuracy of the DIS data. However, for this research it is considered more an issue of the trustability of the data sources, which is addressed in Section 7.8.2.

7.2.1 Issues due to low-detail DIS activity and diagnosis codes

As described in section 6.1.1 most university hospitals register their activities with CBV codes. These CBV codes are more detailed than the ZA codes that are used for declaration and in the DIS database to represent activities. In most cases, the ZA codes do not represent the medical complexity of the operations performed in university hospitals [e4:ha, e7:bi], which directly results in decreased semantic accuracy as the codes do no longer *represent a real world state*. Some quotes stated by the experts during the interviewees relating this issue are provided in Table 10.

Expert	Quote Related to the CBV/ZA coupling issues
e1:ha	*"The activities released by the care authority, NZA, are very general: it does come close to reality but can be much more detailed."*
e1:ha	*"There are also issues with definitions: when can a treatment be called day-care."*
e2:ha	*"We perform certain heart surgeries here for which no classification exists <...> you can book these to "nearby codes""*
e3:ha	*"The possible DBC diagnoses are too limited to express the medical complexity <...> it is badly differentiated"*
e4:ha	*"The couplings do not always correspond to the medical reality: there is a risk which has everything to do with the national couplings"*
e7:bi	*"Less nuances are needed in general hospitals to represent your activities <...> We (UMC) often do innovative operations that do not suit any existing ZA code"*

Table 10: Quotes concerning the semantic issues due to low-detail DIS activity and diagnosis codes

Problems related to the CBV/ZA coupling Technicall, most of the semantic accuracy of the DIS data is lost with the aggregation of the CBV to ZA codes. Section 6.1.1 already describes how this relation between ZA and CBV code is one-to-many: one ZA code could refer to exactly one, or multiple CBV codes. This supports the experts' observation that the ZA code often does not represent the medical complexity: at least, the CBV codes are able to contain more information.

Another possible issue with the aggregation is that no ZA code exists (yet) for the performed activity, which is not uncommon in university hospitals regarding their innovative (and academic) focus [e2:ha]. There is a certain dependency on DBC maintenance at this point: hospitals cannot "just create codes" [e7:bi], these are released by DBC maintenance. Although this situation occurs relatively frequent in university and top clinical hospitals, DBC maintenance has a priority on creating correct codes for the more common (general hospital) operations as those are performed more often [e7:bi].

A final problem concerning the translation of CBV to ZA codes, is that in some hospitals these are confused. For example, normal operative activities in the CBV set all start with the code 33, where the ZA codes start with the code 03. A potential semantic issue occurs with the interpreting of a code such as 33612, which occurred in the hospital of one of the interviewed experts [e5:ha]. This confusion between whether a registered code is a CBV or ZA code is observed in more hospitals[3]. The grouper might reject an unexpected code, but when both codes exist and the abnormality is not recognized by the grouper, the wrong code could be stored in the DIS data [e5:ha].

[3] http://www.dutchhospitaldata.nl/LBZ/Symposium_verrichtingenbestand/Presentaties/Gebruik_CBV_in_een_ziekenhuis.pdf

Adaption of activity codes to obtain a derivable product The grouper ("DOT tree") expects certain activities when a diagnosis has been stated and derives a specific product from this. If these activities are not registered in the treatment sent to the grouper (e.g., because the ZA code does not exist), the product is not accepted (Dutch: *uitvalproduct*), returned to the hospital and cannot be declared. Another possibility is that the grouper leads to a "lighter" product. One of the interviewees provided an example of this: especially in university hospitals it could be the case an innovative operating room (OR) activity is performed which is not recognized (no ZA code exists). In this case the grouper "only" sees the problem (diagnosis) but does not recognize an operating treatment. It could therefore deduce an ambulatory product, which is "lighter" and has a lower reimbursement, while an operating room product was deserved [e2:ha]. If, as in this example, an operation X cannot be combined with a diagnosis Y, the hospital will be inclined to register another operation X' that is similar to obtain a deserved product, rather than not get a reimbursement or receive a lighter product [e5:ha]:

> *"It's a financial system, the output are euro's and not to determine whether a patient has been improved. It is a translation of the actions performed by the specialist with as goal number one to send an invoice."* [e5:ha]

Obviously, semantic accuracy is affected when an activity code is declared and stored in the DIS database, other than the activity that has been performed. The hospitals can also choose *not* to change the activity code, which would eventually influence the population completeness of the DIS data as possibly no product would be deduced and declared. This is further explained in Section 7.7.

Issues due to used diagnoses codes As described, the first issue relates to the high-level of the ZA activity codes required for DIS, where especially information (and semantic accuracy) is lost during the process of aggregating the registered CBV codes to these ZA codes.

Also the diagnosis codes used by the DIS to represent the diagnoses have a similar issue. These *DBC diagnoses* are usually high level and often not medically recognizable [e3:ha]. For each specialism, there is usually only a very limited amount of available DBC diagnoses to register, mapping on many "real" medical diagnoses. Therefore, these DBC diagnoses cannot represent many complex medical diagnoses and a DBC diagnosis stating no more than "other related diagnoses" is not an exception. *"I'd prefer to see the DBC diagnoses removed and have ICD-10 registrations, although this change is not as straight-forward as it may sound"* [e3:ha]. This ICD-10 coding is the International Classification of Diagnoses (ICD), which is used in many countries. With the transition to DOT, these ICD-10 codes are already used by the grouper to deduce a care product. However, the codes are often directly aggregated from the registered DBC diagnosis which results in similar mapping issues and information loss compared to the CBV/ZA code aggregation. Although the hospitals are also able to upload the ICD-10 codes to DIS as well as the DBC diagnosis codes, this is not often done as it only increases the registration burden of the hospitals.

7.2.2 Semantic issues on a sub-trajectory level

The issues resulting in reduced semantic data quality that are described in the previous sub-section, mainly refer to the first phase of the RSAD registration process: the registration itself (Dutch: *Registreren*, R) and related aggregation of activity codes. As described in Section 6.1, after this registration two important succeeding phases are performed. The first is summarizing (Dutch: *Samenvatten*, S) the delivered care by combining the diagnoses and treatments in one sub-trajectory. The second is the deduction of a care product by the grouper based on this sub-trajectory (Dutch: *Afleiden*, A). During these phases, potential loss of semantic accuracy occurs as well. The related issues are not on the level of individual activities and diagnoses, as was the case with the issues described in Section 7.2.1. These issues are on the quality of the combination of these and the care products derived by the grouper.

The main problem is explained well in a quote stated by one of the experts during the interviews: *"It is possible that certain activities are well combined to a DBC, where a combination to another DBC would be preferred as it would result in a care product better relating the provided care"* [e7:bi]. The right combination of the diagnoses and treatments is essential to get right care products, and with it, right reimbursements [e2:ha, e7:bi]. Having a different care product in the DIS data than what was actually performed, does decrease the semantic accuracy of these data.

Differences DBC/DOT This issue could at first be considered as a problem related to the grouper functionality which was introduced with the DOT system. However, the issue was also present in the DBC system: hospitals manually deduced the products they wanted to declare, after which it was validated (RVD model, Section 6.1.5). This validation was to a certain extent similar to the grouper functionality, although performed internal to the hospital.

Actually, the grouper aims to "intercept" some of these semantic issues [e10:dis], as far more semantic "tests" are built in the system than was the case with the RVD model. Recently, the university hospitals information systems' data of 2010 (DBC system) was analyzed by DBC maintenance. This is data more or less comparable to what would eventually get in DIS. In this dataset, 7.2% of all DBCs were removed as the treatment code that would have been deduced by the DOT grouper (simulated) differed to the treatment code that was concluded by the hospitals [14]. These could be cases where the internal validation deduced a poli clinical treatment, where external validation deduced a day treatment [e11:dis].

A drawback is that with this "interception" by the grouper, hospitals have the risk not to derive a declarable product. This could influence the population completeness of the DIS set (Section 7.7). However, with the DOT system, which is an even more activity based costing system than the DBC system, (university) hospitals might choose to change some (activity) codes important for the deduction of the right care product. Not necessarily upcoding or cheating, but merely to "do yourself justice" and receive the deserved reimbursement [e1:ha, e5:ha]. This was also observed in the previous sub-section. Just like with the old RVD registration model, the added grouper functionality with DOT can therefore also be a reason for a decreased semantic accuracy of the DIS data.

7.3 Time-Related: Currency

According to Batini and Scannapiena [7], Currency concerns how promptly data are updated. In the context of this research, this refers to the time between the time of providing the care (change in real-world state) and the registration of these activities and diagnoses internal to the hospital. Although currency could technically also relate to the time between the activities/diagnoses and the update of the DIS tables, the choice is made to take the first option. The latter is described as the timeliness sub-dimension, which is further explained in Section 7.4.

As can be seen in Table 7 (page 65), currency is graded as one of the lowest (worst) quality dimensions although the expert opinions differ largely ($\mu = 6.5$; $\sigma = 1.8$). For this dimension, the experts used the currency of their own internal registrations as reference. The issues provided by the experts are roughly the same, only their interpretations on the severeness of the issues differ. Table 11 shows some relevant quotes concerning currency.

Expert	Quotes Related to Currency
e1:ha	*"The activities are usually quickly registered, within a week I'd say. Opening a DBC is automatically performed, but the coupling of the right diagnosis to the DBC takes a lot of time and effort"*
e3:ha	*"The currency depends on the specialism and on the discipline of the specialist. As long as he does not register the diagnosis, it cannot be combined in the DBC."*
e4:ha	*"Usually the currency is good. <...> The problem is with activities on emergency care, which sometimes takes a week before it is put in the system"*
e5:ha	*"Usually, there are sub systems. For example, a lab system. A labrequest is performed and executed, but it can take up to a week before it is in the system. I think that is too long, it should be one day (OLTP system)."*

Table 11: Interview Quotes concerning Currency

There are different aspects of the treatment that are registered. Usually, with regard to the opening of a DBC (with nothing attached) there are hardly any problems. This is often performed directly with the first patient contact [e7:bi], manually at the patient visit, or, if a specialist does not open a DBC, it is even automatically performed [e1:ha].

Currency problems related to the diagnoses However, determining the right diagnosis code and coupling it to the right DBC usually takes a lot of time. This time differs, depending on the discipline of the specialist and on the specialism. Usually the "viewers" (e.g., internist) are slower than the "cutters" (e.g., surgeon) [e3:ha]. The currency of registering and coupling the diagnoses can take up to a week. The same currency exists for the registration and coupling of certain activities. Emphasizing this, one of the interviewees recommended the following:

> *"These data should be in the DBC system much faster. By increasing the communication speed between the different systems. By having more tight procedures like always registering the agenda contents at the end of a consultation hour. Instead of saying we wait a week with this, because there are always corrections and such. That is only because the process is not streamlined."* [e5:ha]

7.3.1 "Floating Activities" (Dutch: *Zwevende verrichtingen*)

Although there are usually not many problems concerning the opening of a DBC, there are occasions where activities cannot be attached to a DBC because the DBC is not created yet (e.g., a lab activity is performed before the patient visits the hospital), or there is an erroneous DBC (e.g. with a wrong data or specialism). These issues are returned to the specialist for feedback [e1:ha, e5:ha].

Until the specialist corrects the DBC issues, or the (correct) DBC is created, these activities will be "floating" (Dutch: *Zwevende verrichtingen*). All hospitals visited for this research were dealing with these floating activities, usually a set of activities that were solved within one week, replaced by a new set of comparable size. One of the experts describes an important issue causing this phenomenon:

> *"With every care request, the "requester" is provided, which is directly registered. There is no control whether this requester has an existing DBC. That is the problem. The whole workflow management system, the communication of orders, is absent in most hospitals. Therefore, everything is based on a transactional system of registering a care request together with the requester. This problem is everywhere and one of the largest causes of floating DBCs"* [e5:ha].

The amount of the floating activities differ, one (general) hospital had a constant "bucket" of 2.000 floating activities on hundreds of thousands activities (more than one hundred thousand first consults) [e5:ha]. Another (university) hospital has 2,2 million activities, with 75.000 floating activities (3,4%, usually all solved within a week) [e7:bi].

Within one hospital, the total amount of floating activities is usually stabile. Thus, at the end of the registration process when the data is declared and submitted to the DIS, most of these floating activities are combined to a DBC [e5:ha]. However, in some rare cases, the activities "float" for longer than a week. *"There have been situations where we combined floating activities to DBCs more than a year behind."* [e1:ha]. Depending on the arrangements with the insurers, hospitals are usually able to couple and declare those activities. But especially when the activities are not relevant for the derived (and declared) care product, these are never coupled, and removed [e1:ha]. Those cases influence population completeness of the DIS data, further explained in Section 7.7.

With the DOT system, process improvements have been initiated in certain hospitals to improve the awareness of the importance of fast and good registrations [e1:ha, e5:ha]. This is because of the financial changes, and because of the automatic closing: if a path has to be closed after 42 days, a large risk of un-declarable activities exists when these have a currency of one week [e7:ha].

7.4 Time-Related: Timeliness

The timeliness dimension, which expresses how current data are for the task at hand [7], is considered a contextual quality aspect. However, the age of the DIS data (not related to the task at hand) is context independent and is (thus) described in this section. It is important to realize that the DIS contains only those DBCs that are declared at the insurers (see Section 6). This implies that tariffs have to be negotiated, the trajectory (care product) has to be closed and validated or deduced by the grouper, and it should have been declared. After that, the dataset has to be prepared for DIS submission.

The aspects are described below and conclude to a usual (timeliness) delay of at least a year and one month. DIS data currently (May 2012) includes data on DBCs closed up to December 2010.

Tariff Negotiations As long as reimbursement tariffs have not been negotiated with the insurers, no invoices can be sent which implies the data cannot be submitted to DIS yet [e6:bi]. This dependence on health insurers can be up to one year, e.g., there are still no tariffs of 2012 [e3:ha, stated in May 2012].

DBC Closure One of the main aspects influencing the timeliness of the DIS data is the fact that only closed trajectories are submitted. This closure is inherent to the system and differs in the DBC and DOT systems.

- **DBC**: In the DBC system specialists were allowed to manually close a path, with a maximum duration of one year.
- **DOT**: In the DOT system, the paths are automatically closed after a predefined timespan. This timespan cannot be closed earlier, unless in certain exception situations (e.g. the patient dies). The timespan differs, based on the activities within the trajectory: DBCs following an earlier DBC (Dutch: *vervolg DBCs*) are always (exactly) one year; a primary DBC is 42 days after a clinical intake (Dutch: *opname*), otherwise 90 days, although some exceptions exist (e.g., gynecology) [e3:ha].

The timeliness will increase with DOT. Although the fastest duration of a product is still one day (in case of the death of a patient), because of the predefined closure rules, on average it will take longer before a product is closed [e5:ha].

Furhter validation and declaration One interviewee mentioned often further internal validation was required after the closure of a trajectory, which could easily take a long time in case of errors. Furthermore, declaration also takes time: the combination of the two can take another one or two quarters of a year [e4:ha].

DIS submission schedule The DIS submissions are usually scheduled once every one to three months, depending on the hospital. This increases the delivery time with at least one more month. However, most hospitals perform the validation and declaration within this month (or three-month timespan) [e10:dis].

Delay of hospitals On top of the timeliness issues mentioned above, some hospitals are months (one case of six months was found) behind their DIS delivery [e10:dis, e5:ha].

Time between DIS delivery and database entry When DIS receives the data, it is immediately processed and placed in the database within one day [e9:dis].

7.5 Time-Related: Volatility

Volatility characterizes the frequency with which data vary in time [7]. In this context, this refers to the frequency updates of already stored DBCs are submitted to the DIS database. Table 7 (page 65) shows the grades provided on this dimension. As can be seen, two interviewees graded this dimension somewhat low (6.0, 6.0), while three graded it provided it a higher grade (8.0, 8.0, 9.0). With an average grade of 7.4 ($\sigma = 1.3$), this volatility was not considered to be of large and significant influence to the overall database quality, neither positive nor negative. There are no real issues identified: merely the aspects influencing volatility, which are intrinsic to the system, are identified.

Hospitals are able to submit changes to an already submitted DBC, months or years later. DIS still receives (updated) DBCs that have been closed in 2007 or 2008 and are already in the database [e9:dis, e10:dis]. These updates are very common: in one hospital, 30% of the DBCs were eventually revised (and updated in the DIS set) [e5:ha]. Therefore, the volatility of the DIS database can be considered ”high”.

One frequently occurring update, in the DBC system, was a change in the end-date of the DBC. As trajectories could be closed manually, sometimes an (unexpected) activity was performed after the closure. In those cases the end-date of the DBC was changed to the date of the newer activity so it would be included. This resulted in an update to the DIS (and insurers) in case the DBC was already submitted (not unlikely as it was already closed). Although the height of the reimbursement might not always increase with this category of changes, hospitals have to update this information to the insurers: their administration should be the same compared to the hospital administration as the insurer eventually also verifies the data [e5:ha].

Updates from Insurers These insurers can verify the data up to five years back, forcing the hospitals to change their data [e3:ha]. Most interviewees state that ”their” hospital also submits these updates to DIS [e1:ha, e4:ha, e7:bi, e9:dis]. One of the experts stated that DIS only accepts data changes on data up to a certain age and they therefore only submit older data to a certain limit [e3:ha]. This means that some later changes (especially insurer updates) are not submitted to DIS anymore and the data are not consistent to what is actually reimbursed. One of the experts even seriously doubted all hospitals would submit such old (insurer) data updates [e5:bi], and some tests on this should actually be performed:

> *”The question is if that, due to the age, leads to a change in the DIS delivery. <...> You would want to see that the hospital deliveries to DIS also contain corrections of older years. Because it would be really strange if a hospital never submits something older than two years. All would have been submitted good at once, and trust me, that is not the case”* [e5:bi]

Update method in DIS When confronted with an update of an already stored DBC, the DIS does not remove the older version. Merely, it flags it as ”expired” and stores the new one: DBCs are only functionally, not physically [e10:dis]. Furthermore, with the DOT automatic closure rules, situations where later performed activities are updated to the DBC afterwards are less likely to occur, which would decrease the volatility of the DIS set [e5:ha]. However, updates will still occur as other examples are also imaginable, e.g. when the health insurer of a patient has changed [e4:ha].

7.6 Completeness: Column Completeness

Column completeness is defined as a measure of the missing values for a specific property or column in a table. Table 7 (page 65) shows the gradings provided to this dimension. Although these grades are all positive ($\mu = 7.5$; $\sigma = 1.2$), two of the experts graded clearly lower (6.0, 6.0) than the other three (8.0, 8.0, 9.0). This was caused as not all experts considered *all* fields in DIS (required *and* non-required). The differences between these views and the results on the grading are described below.

Expert	Quotes Related to Column Completeness
e9:dis	*"Some of the fields are required: these are certainly all in the DIS set, this is validated before entry. However, also non-required fields exist"*
e3:ha	*"We submit what is required, we do not submit what isn't"*
e7:bi	*"We submit all required fields. Regarding the facultative fields, we only submit those items we can easily generate"*
e4:ha	*"Required fields are required in the whole registration process, those have to be filled <...> The facultative fields are, with high probability, not filled."*
e5:ha	*"The required fields are validated so should be in. <...> But you have to be careful when you want to utilize facultative fields in a research"*

Table 12: Interview Quotes concerning Column Completeness

The quotes in Table 12 summarize the (shared) opinion of all interviewees: all required fields are certainly in the DIS data. This is validated before the database entry, if these fields are missing the entire product is not included in the DIS, which would influence population completeness (Section 7.7) instead of column completeness. Therefore, there can be concluded there are no (column completeness) issues with regard to the required fields in the DIS data.

More interesting are the non-required, or facultative, fields, where the main issues are. Either, none of these fields are submitted, or only those fields that can easily be generated from the existing hospital data are. One of the interviewees added:

> *"This relates to the academic culture: if something is not mandatory, specialists will rarely put effort in registering it."* [e4:ha].

The same expert further provided a possible recommendation with regard to these required fields. This statement also relates back to observations made in Section 7.2.1 on the high-level of the DBC diagnoses, where ICD-10 diagnoses are already possible in DOT, but hardly used by the hospitals yet:

> *"Possibly, DIS should have more fields required. If you want to prepare for mandatory ICD-10 registrations, you have to require that ICD-10 field already this year or the hospital will not start working on the delivery of it."* [e4:ha]

7.7 Completeness: Population Completeness

Population completeness evaluates missing values with respect to a reference population [7]. Based on the grades provided by the experts (Table 7 on page 65), it is together with semantic accuracy considered as the largest quality risk of the DIS data ($\mu = 6.1$; $\sigma = 1.5$). All experts recognized the same population completeness issues; discrepancies in the gradings were caused by the perceived impact of these issues on the total dataset (and reliability of analyses). An interesting observation is that the interviewee providing the highest grade (e5:ha) was actually attached to a regular hospital, whereas the other interviewees were employed in university hospitals.

Population completeness has to be assessed with respect to a reference population. Concerning this, two levels can be differentiated. The first is the level of treated patients and the DBCs (or care products) they received. The second is more detailed and answers the question whether, besides all treatments, all performed activities in those treatments can be found in the data. The experts generally agree that all patients that have been treated in the hospital are stored in the DIS database: both the patient and his delivered care product, regardless of the semantic accuracy of this product. However, more missing entities exist when looking at the treatments in more detail: all performed activities. Two main issues exist with regard to (non-) storing of activities: either these are not recognized and rejected by the grouper (Section 7.7.1), or these are irrelevant (Section 7.7.2). Table 13 describes some general quotes experts stated on this quality dimension.

Expert	Quotes Related to Population Completeness
e4:ha	*"I think this is the biggest issue, it is not good. Sometimes things are forgotten, especially activities, or they are not placed in the right trajectory or disappear in the system and stay only in the internal database. DIS receives none of those"*
e5:ha	*"Eventually you will miss a couple of percentages. On rejections, or corrections from insurers that are not submitted"*
e5:ha	*"You end up with like 98,5% of the data that passes all validations and which is used for the financing of the hospitals. Then, you can hardly keep saying that the data is hopeless. I do not believe that"*
e7:bi	*"Patients and DBCs are very good (all exist in DIS). But if you look at the combinations of patient-carepath-subtrajectory-activity: everything is in what should be in. But you do miss data especially the very specific activities such as those from neurosurgery. <...> This does not occur often but these are important, e.g. activities with which we can justify our academic setting"*
e10:dis	*"If you are talking about the completeness of the data you can differentiate two levels: DBCs and activities. On DBC level we will score pretty high. This has externally been assessed once and it was a good match. <...> On detail level, really the activities, you can't really say something about it, it is differentiated and depends on how well a hospital has organized his own administration"*

Table 13: Interview Quotes concerning Population Completeness

7.7.1 Invalid activity codes and care products

As described in Sections 6.1.1 and 6.1.3, problems exist with the aggregation of registered activities (CBV codes) to national activity (ZA) codes and the grouper rejections. This can result in the hospital changing the ”bad” activity code to a code that resembles more or less the same and which is accepted by the grouper, influencing semantic accuracy (see Section 7.2). However, another scenario is that hospitals do not change the activity code and the activity is not used for the care product deduction by the grouper. In those cases, the activities are never stored in the DIS database, influencing the population completeness. Although hospitals will in those cases receive either a ”lighter” product, or no reimbursement at all, with the old budgets still existing in the DBC system many university hospitals chose this approach [e1:ha]. That this scenario occurs often is, besides through the interview results, supported by a technical report on the 2010 dataset recently written by DBC Maintenance. This report observes that data from the university hospitals' information systems (gathered by the NFU) includes more activities in one care product than the DIS data of the same university hospitals [14]. With DOT, and the performance reimbursements (all budgeting pillars are removed), the completeness issues due to these invalid activities will probably be less eminent, as hospitals will need the reimbursements [e1:ha].

Relevant quotes describing this phenomena of invalid activity codes are stated in Table 14. It is interesting to see the discrepancies in opinion between the different stakeholder positions: these quotes regard 3 hospital administration experts from a university hospital, one hospital administration expert from a general hospital (e5:ha) and one employee DIS (e10:dis).

Expert	Quotes Related to Invalid activities/DBCs
e1:ha	*”You sometimes notice products are not valid due to the complexity of the treatments in the UMCs. We deliberately decided not to complete these products by adding an acceptable activity code. At least, up to now. But this might just change if we are not reimbursed for those treatments anymore. Before <DBC system> it did not matter: we got the money anyway.”*
e2:ha	*”Only everything once declared, is known to DIS. What does not fit the national product structure is not visible. But exists. Only, it does not get further than the validation or grouper phase.”*
e3:ha	*”Often it is the case that the completeness issues relate to the difficult patients. Those, that do not fit the DBC patterns. These are rejected <by the grouper, Dutch: uitvallen>, but are very interesting”*
e5:ha	*”This <missing activities, e.g. due to rejections> might be the case, but if you look at relative missing entities it is only one or two percent. Because something is eventually made up for these activities anyway.”*
e10:dis	*”This is mainly the case in university hospitals. Those perform relatively more activities that are not defined, so these have more rejections”*

Table 14: Interview Quotes concerning Missing treatment combinations/products

The incompleteness on activity level has a large influence on the hospitals. Cost prizes are determined based on DIS data (the set sent to DBC maintenance). If activities cannot pass the grouper, these are not in the DIS set and cost prizes are calculated that are too low, especially for academic and top clinical hospitals [e4:ha]. Therefore, starting in 2013, DIS will also request the input codes connected to a trajectory [e2:ha]. However, this only increases the information received on the declared products and does not solve the primary issue that products that cannot be declared due to these issues are never in DIS [e2:ha].

Recommendations based on the interviews The issues described in this section, on activities and products DIS never receives due to the aggregations and grouper functionality, are realized both by DIS and by the hospitals. An interesting observation based on the interviews, is that both parties state a lot would be solved if DIS received all delivered care, even those that (with DOT) could not be deduced to a declarable product by the grouper. Some quotes really stating this, are provided in Table 15.

Expert	Quotes stating DIS should also receive invalid activity codes and care products
e10:dis	*"We never see this! It is an observation which can only be made by a hospital, not by DIS."*
e10:dis	*"Actually, we once offered them to send all those activities, to see if we could do something with it. It is easy to say that one has so much or so few, or how much underivable products (Dutch:* uitval*) there are. But you are only sure once you start measuring."*
e9:dis	*"For us, it is very interesting that the hospital says this. Because, for us it is not rejected (Dutch:* uitval*) because it's not in our scope up to this moment. But it would be nice if it would become in our scope, so we could recognize and call it as* uitval*."*
e5:ha	*"DBC Maintenance currently receives signals of all kind of things that due to issues and innovations are not deduced correctly by the grouper. With this, they start working on it with different stakeholders. But they never do that based on data of rejections (Dutch:* uitval*). It is always based on incidents. While you could also say like, we are busy three months and have three months of data, we will analyze everything and see what goes wrong"*
e5:ha	*"So actually, DIS should say, we do not mind we want everything. Even DBCs that are not closed, also DBCs that were not deduced to a valid product, even the floating activities."*

Table 15: Quotes stating DIS should also receive invalid activity codes and care products

7.7.2 Activities Irrelevant to the declaration

Another issue with regard to the population completeness on activity level, again relates to the financial incentive of the DIS data (and the DBC/DOT systems). The DIS data stores all treatment data that are financially relevant: those diagnoses and activities that have been provided to the grouper in order to receive a reimbursement. Especially with the DOT system (automatic deduction of care products, based on the presence and combination of certain diagnoses and activities) this directly influences the population completeness of the activities. The hospitals have no direct incentive and need to register those activities that are irrelevant to the grouper deduction (and thus the money they receive). As the previous section (7.6) described, hospitals and specialists are inclined not to perform (administrative) activities that are not mandatory. This explains why these "irrelevant" activities are likely to miss in the DIS data; lab activities are a good example of this. Some illustrative expert quotes related to this, are provided in Table 16.

Expert	Quotes Related to Irrelevant activities
e2:ha	*"Billed (Dutch: gefactureerd) is full-continuous the magic word"*
e1:ha	<Related to long-term floating activities:> *"If it does not influence the DBC, for example a lab activity does nothing, you will not insert them anymore. But then these are eventually not in DIS so your profile is not entirely complete"*
e2:ha	*"I can even approach it minimally: the only activities i register are the typical activities (Dutch: typerende verrichtingen) I need to have a declarable product, I drop the rest"*
e10:ha	*"The problem is mainly with the delivered activities. One DBC should contain multiple activities. We know that this is incomplete for certain hospitals and not as it should be."*

Table 16: Interview Quotes concerning Irrelevant activities

This also translates to the organization of the activity registration: usually a list of activities a specialist can register (a choice in a drop-down menu) exists: this is potentially incomplete. Such listings do (obviously) include all activities relevant to the deduction of right care products and thus to the declaration. However, if an activity is performed and the specialist cannot register it in his system, it will not be registered at all: usually without further notion as these activities are often irrelevant to the declaration [e4:ha].

Recommendations based on the interviews One of the experts (e5:ha) explained that the DIS data is currently only validated on data that is delivered. He states that there should also be validations whether hospitals have not made specific organizational choices to consequently not deliver certain data. Thus, not really on the level of one operation ("are the last 10.000 lab activities present), but on the level of all care products ("are lab activities for this product consequently not uploaded, while you would expect these activities").

> *"You would want to state: I expect at least these types of activities, approximately in those amounts. This is a validation you could do before performing analyses. In all research you perform on these data, you will experience that you need to check for completeness."*

7.8 Trustability of the Data Sources

The last three intrinsic quality dimensions provided by Wang and Strong are *Believability*, *Reputation* and *Objectivity* [85]. Believability considers whether data provided by a certain source can be regarded as true, real and credible; reputation considers the trustability (credibility) of the information *source*; and objectivity refers to the extent to which the sources and provided data are unbiased (not subjective) [7, 85].

There is a strong relation between these three dimensions: all relate to the trustability of the data as a consequence of the sources of the data. Although this has been observed earlier by Batini and Scannapieca [7, page 35], the three dimensions were referred to separately during the interviews. However, during the interviews, the three sub-dimensions were often confused and interchanged, which also shows from Table 17 (of the twelve grades, for every expert the results were graded very similar, with two exceptions).

Quality Dimension	e1:ha	e3:ha	e5:ha	e7:bi	Average	St. Dev
Believability	6,0	8,5	9,0	7,0	7,6	1,4
Objectivity	6,0	5,0	9,0	9,0	7,3	2,1
Reputation	6,0	8,0	9,0	8,5	7,9	1,3

Table 17: Grades of the Trustability related Quality Dimensions

Most information gathered after raising these sub-dimensions during the interviews were firstly related to the identification of the exact data sources. These results are discussed in Section 7.8.1. Secondly, a conclusion on the data provided by these sources was stated, based on a mix of the reputation, objectivity and believability. Batini and Scannapieca refer to this as ”the quality of the data sources” [7, page 35], or ”trustability of the data sources”. An analysis of the interview results on this is provided in Section 7.8.2.

7.8.1 Sources of Data Entry

Diagnoses When the DBC system was introduced, it was intended that the specialists would enter the DBC data, and some hospitals or departments are still organized this way [e7:bi]. However, in most cases the specialists only perform the registration of the (DBC) diagnosis [e1:ha, e2:ha, e5:ha, e7:bi]. There is currently one Dutch university hospital for which the specialists are not involved at all, but where the complete DBCs are deduced from the basic registration (VUmc). In this case however, the specialist still has to provide a diagnosis (the deduction has to be based on something) but this is performed in a more detailed registration.

Activities In the cases the specialist only registers the diagnosis, the registration of all other aspects (activities) are not performed by himself, although he (she) is still responsible for it [e2:ha]. These aspects are then performed by administrative personnel or directly by an information system. For example in some hospitals, the agenda of the doctor is at the end of the day processed (by administrative personnel) and updated with the patients that were seen but not scheduled and those that did not turn up. After this processing step, the appointments are automatically translated to (first or subsequent) consultation activities (Dutch: eerste of vervolg consults) and combined to the corresponding DBC [e5:ha,

e7:bi]. Likewise, care days (Dutch: verpleegdagen) can automatically be deduced from admission registrations [e5:ha], which is also entered by administrative personnel. For activities performed at external departments, such as the lab or radiography department, codes are attached in those systems which are provided to the DBC system(s) [e5:ha]. Another possibility is that the hospital administration (or even the specialist himself) fills checklists of what has been performed by a patient, which is then translated to a DBC (ZA) activity code [e7:bi].

Opening and Closing of the DBC With the DOT system, closing of a DBC is performed automatically following closure rules (Dutch: afsluitregels). However, the opening of a DBC still has to be performed. In some hospitals, this opening is purely an administrative activity which is performed by admistrative personnel such as a secretary [e2:ha, e15:con], or performed when a diagnosis is found in the system [e3:ha]; some hospitals combine the opening of a DBC with the attachment of the diagnosis and as such, this is an activity performed by the specialist [e5:ha].

The aforementioned stakeholders who are responsible for the original data entry are considered as the best possible stakeholders considering the possibilities. In the ideal situation, from an administrative point of view, all registration are performed by the specialists. However, regarding their primary responsibilities and limited time, it is good some tasks are performed automatically or by administrative personnel such as a secretary, although these stakeholders have the drawback of being less involved in the care process [e.g. e1:ha].

7.8.2 Quality of the Data Sources

Most experts do have some concerns related to trustability of the data, with relation to the data sources. However, they emphasize that discrepancies of the data from medical reality are not necessarily caused by malevolent intentions, but due to the underlying system [e1:ha, e4:ha, e5:ha, e7:bi]. For example, there are cases where specialists might not register the right corresponding activity code but a ”near hit” in order to get the right product (see Section 7.2). According to the specialists this is not performed intentionally to receive undeserved more fundings (upcoding). They do however point out that this especially relates to academic hospitals where specialists are in salaried employment and do not ”feel” this upcoding directly in their benefits. Specialists in the regular hospitals are self-employed and share the benefits of the hospital, thus might have more such incentive [e1:ha, e4:ha]. This is also supported by the PhD dissertation of Hasaart [40], which is an exploratory research on upcoding in Dutch hospitals.

Most of the interviewees provided comparable examples related to the ”trustability” of the data (sources) and to the semantic accuracy. This is not a surprising result as the definitions of the dimensions largely overlap. Furthermore, there is a cause-consequence relation between the two: when a data source is less credible of delivering the right data, this will show in (possibly amongst others) the semantic accuracy of these data. Specifically regarding the objectivity of the data, one expert stated the concern that the system in itself is subjective, where instructions are not always clear and definitions are multi interpretable [e3:ha]. This explains the one lower grade provided by that expert (Table 17).

Concluding, it can be stated that the experts consider the stakeholders responsible for the data entry suitable, or at least ”the best possible option”. The issues related to the believability, objectivity and reputation of the data (sources) are mainly caused by issues in the system itself, which have been identified earlier in this section.

7.9 Conclusion of the Intrinsic Data Quality Analysis

This section has described the analysis of the expert interviews on the intrinsic data quality of DIS. Of the nineteen performed interviews, nine have been specifically on this data quality dimension. The other interviews were specifically aimed on determining the contextual data quality of DIS to evaluate efficiency or patient safety. However, during all interviews at least some attention was to the intrinsic data quality. The interviews, and this section, were structured by the different intrinsic quality dimensions constructed in the data quality framework. This framework is shown in Figure 5 (page 31). Below, the leading opinions of the different experts for the quality dimensions are described.

7.9.1 Accuracy

Regarding accuracy, syntactic and semantic accuracy are recognized. Syntactic accuracy refers to the extent to which fields in the database represent *possible* real world states (e.g., is the field containing ”activity code” always filled with an existing activity code). Semantic accuracy refers to whether this activity code represents the *right* real world state (thus, does the code represent what was actually performed).

Regarding the *syntactic accuracy*, there are no real threats. Most of the fields in DIS are standardized and based on external tables. Before inclusion in the DIS database, this linking to the external tables is validated. If invalid, the related products are not included in the database.

However, the DIS data is expected by the interviewees to have most issues on *semantic accuracy*. There are two main issues identified for this. First, the codes used in the DIS database to represent diagnoses and especially activities are not detailed enough to represent the medical reality. Especially university hospitals use another coding to internally register their performed activities. When aggregating these activities to the national activity codes (ZA codes) used for DIS, often a lot of information is lost and there is a risk of a semantically incorrect aggregation. Second, diagnoses and activities are not always coupled correctly after the registration, which results in a semantically incorrect (by the grouper deduced) care product.

7.9.2 Time-related

There are two intrinsic time-related dimensions: currency and volatility. Furthermore, the timeliness (age) of the data was assessed during the interviews as well, although mostly contextual (whether the age is good enough for the specific task).

Currency regards to the time between performing the care and the internal registration of this. In many cases, determining the right activity or diagnosis (code), and coupling both to the right DBC, usually takes a lot of time. This can take up to a week. When hospitals cannot couple an activity to a DBC yet, these are referred to as *floating activities*. In all hospitals visited, there is a significant amount of floating activities, although the size of this set is usually stable: most floating activities are coupled after one week, where a comparable amount of new floating activities have been originated.

The time between registration of the provided care and inclusion in the DIS database (*timeliness*) is at least a year and a month. This mostly depends on the fact that only closed paths can be submitted to DIS, which can easily take up to a year. Furthermore, hospitals update their DIS data monthly.

Regarding *Volatility*, there are some changes to the data after submission, although not significant. As only closed products are send, from within hospitals mostly not much changes (especially not with the DOT system). However, health insurers also provide tests of the data, sometimes only five years after the first submission. If they conclude some product is not valid (e.g., wrong coupling of activities and diagnoses), this has to be updated in the hospital's system, also resulting in an update of the DIS data.

7.9.3 Completeness

For the intrinsic completeness dimensions, column and population completeness are identified.

Regarding *column completeness*, whether all fields are filled (not NULL) this is true for all required fields. As with syntactic accuracy this is also validated on. However, for the non-required fields, this will usually not be the case. This relates to the culture within the hospitals: if something is not mandatory, specialists will rarely put effort in registering it.

Population Completeness regards to missing values with respect to a reference population. All patients that have received care in the hospital will eventually be found in DIS, with a related care product. Thus, on that level there are no population completeness issues. However, whether all these care products contain all performed activities, is highly unlikely. First of all, when aggregating the activities as the hospital register it internally (CBV codes for university hospitals) to the national activity codes used in the DIS database (ZA codes), there is not always a valid link. Thus, a CBV code exists, but this might be such a new and innovative activity that no national ZA code exists for it yet. Hospitals can in those cases choose not to register this activity for DIS. Furthermore, with the grouper functionality of DOT, not all activities are relevant to the deduction of a care product. For example, in many cases the same product is derived whether or not lab-activities are registered. Hospitals often choose not to register such activities ("if something is not mandatory, hospitals will rarely put effort in registering them").

7.9.4 Trustability

Regarding trustability of the data, sources of the data entry have been identified and whether or not these sources (employees) are most suitable for the task. The specialists usually only register the diagnoses. The performed activities are either performed by administrative personnel or by an information system (e.g., by automatically deducing activities from a specialist's agenda). Although ideally all registrations are performed by the person most knowledgeable on the provided care, thus the specialists, taking time restrictions and priorities into account the aforementioned stakeholders are considered most suited for the registration tasks. There are some concerns of the experts on the trustability of the data with relation to these data sources. However, they emphasize that discrepancies of the data from medical reality are not necessarily caused by malevolent intentions, but due to the underlying system.

8 The Contextual Data Quality of DIS to Evaluate Efficiency

The previous section provided an analysis of the intrinsic DIS data quality based on the data gathered during the expert interviews. With it, information is assembled to answer the first research sub question of this thesis research (which is performed in the conclusion in Section 10). This section provides an analysis of the related interview data in order to answer the second research question: *"What is the contextual data quality of DIS to evaluate efficiency in Dutch hospitals and how is this determined by intrinsic data quality of DIS?"*.

Recall that contextual data quality refers to the "fitness for use" or "usefulness" of the data in a specific context [57, 74, 85]. In the context of efficiency assessed in this research, this relates to the usefulness of the data to perform specific efficiency measurements or research (see also Sections 4.1 and 4.2.5). Therefore, based on the interviews, first more insights into the currently used efficiency measurements are acquired. This is described in Section 8.1. Succeeding, Section 8.2 describes the available data in the DIS data that are potentially useful for these categories of measurements. Section 8.3 then describes the limitations of these data with respect to efficiency measurements and research. With this knowledge on the context, the available data and the limitations of the data, Section 8.4 then describes the possibilities of the data for efficiency research. The combination of the limitations and possibilities should provide more understanding of the "fitness for use" of these data in the context of efficiency research.

8.1 Efficiency Measurements from Interviews

As stated, one of the first goals of the efficiency interviews was to assess current methods used to measure efficiency, known or used by these interviewees. Two levels of efficiency measurements have been identified: interhospital, which compares hospital-wide efficiency scores of different hospitals; and intrahospital, which assesses efficiency in more detail within a hospital. The first is explained in Section 8.1.2; the second in Section 8.1.3.

As explained in Section 4.2, different levels of granulation to measure efficiency exist: technical, allocative and economic efficiency. Before understanding the specific hospital efficiency researchers, knowledge on how these domain-independent concepts translate to this specific (hospital, DBC) setting is required. Therefore, this is explained first, in Section 8.1.1.

8.1.1 Technical, Allocative and Economic Efficiency in the context of this research

Section 4.2 provides a detailed description of the general definition and differences of these concepts. This paragraph describes the definition and application in this specific research (domain). Explaining these different levels of granularity of efficiency research will also aid in describing the possibilities of the DIS data in relation to this level of granularity.

Technical Efficiency describes the ratio between the amounts of input and amounts of output. In this context, this relates to the total of (different) activities (Dutch: *verrichtingen*) performed to receive the same care product: less activities would be more efficient (disregarding patient safety issues). There are different activities involved for one DBC, so two different approaches to look on technical efficiency. First, the totals of the different activities can simply be counted where more activities is technically inefficient. The second approach, Free Disposable Hull, is more restricted. In this approach *"You only*

qualify a hospital better than another one if it delivers (at least) the same production but uses less of all *inputs"* [e16:sci].

Allocative Efficiency refers to the used mix of different input factors to produce an output. In this context, this relates to the optimal proportion of different activities used for the same treatment. For example, whether less costs are involved to perform a cheaper operation which requires the patient to stay more days in the hospital (Dutch: *ligdagen*) which also costs money, or to perform a more expensive operation after which the patient can immediately be released. This approach requires information on the costs and on the prices of the different activities [e16:sci].

Economic Efficiency is defined as the product between technical and allocative efficiency where a maximum output is achieved with an optimal combination of input factors. In this context, where the output is predefined as one delivered care product, only technical (total amount of activities) and allocative (total costs of activities) are applicable.

8.1.2 Hospital-wide efficiency measurements

The focus of most scientific publications on hospital efficiency is on hospital-wide measurements. These compare hospitals as a whole, where the total of inputs and outputs are hospital-wide determined (although possibly accumulated from different specialisms or departments). Section 4.2 describes this approach and corresponding techniques (e.g., DEA or SFA) and level of granularity (technical, allocative, economic efficiency) in more detail.

There are two important notions to make which will be important to determine the (im)possibilities of the DIS data (with knowledge on the precise contents of the data). These are the input and output factors often used in literature and further explained by the two interviewees closely related to this category of research [e16:sci, e17:sci]:

1. As input factor, costs is generally used: the total costs made by a specific hospital in a specific year. This is often decomposed into multiple dimensions: staff, capital and materials, and sometimes even further decomposed (e.g., different categories of staff).

2. The output factor used in these researches is usually the production, although different variations exist. Determining this production (or output) is notoriously difficult (also described in Section 4.2). In most cases, the total amount of admissions (or a combination of released and deceased patients) is used to measure production.

8.1.3 Efficiency Measurements within hospitals

Two efficiency measurement techniques have been explained during the interviews, which are used within hospitals to compare performances of different departments, specialisms, treatments or specialists [e7:bi, e8:bi]. These techniques are mainly based on internal (DBC!) registrations of the hospitals.

The first method is related to the current political opinion of the specialization of the hospitals in The Netherlands, which aims on reducing the different treatments performed in one hospital and have hospitals specialized in certain treatments. In this approach, multiple DBCs are categorized into a limited amount of "attention fields", which are medically more recognizable than DBCs [e7:bi]. Efficiency is measured as the difference (percentage) between the reimbursement of the product by an insurer and the total costs made for the treatment: the degree of coverage (Dutch: *dekkingsgraad*). This

is mapped against the relevancy of the treatment to this specific hospital, based on two quantitative and two qualitative measurements. This enables the hospital to quickly visualize the attention fields that are neither (cost) efficient and not really relevant to the hospital, which could help in the possible specialization process [e7:bi].

The second method compares the cost prices for a specific DBC of that hospital to the average national cost prices (of 22 hospitals). This differs from the other hospital which compares the cost prices of a (group of) DBC(s) to the insurers reimbursement. The differences in cost prices are mapped against differences in profile: *"How many activities are on average performed for a specific DBC"* [e8:bi], compared to a national average as well. DBC maintenance released this average amount of the different activities related to one DBC. This approach is not designed for specialization purposes, but for costs savings purposes: the savings potential of a DBC is also visualized in this matrix mapping differences in profile and costs [e8:bi].

It can be observed that the (two)efficiency measurements within hospitals are not one dimensional: both techniques are visualized in a (two dimensional) matrix. A difference in cost prices is used in both approaches, one compared to the reimbursement and one mapped against the average of approximately twenty other hospitals. One hospital maps this against medical relevancy and one hospital maps this against another efficiency measurement: differences in profile.

8.2 Contents of the DIS data related to efficiency research

Two tables in DIS contain information relevant to efficiency calculations: these are the table containing information on the sub-trajectory (DBC) and the table containing information on the delivered care profile: the performed ZA activities. The most relevant fields are numerated below:

- Sub-Trajectory
  - **DBC Hospital Costs** Costs made by the hospital to deliver the specific sub-trajectory
  - **Honorarium Total** The total declared amount of honorarium.
  - **Declaration Code** Code referring to external "tariffs" table which contains information on reimbursement amounts for the specific sub-trajectory
  - **Claim ZVW** For specific products, indicates whether the health insurer had to reimburse the treatment, or the patient for himself
- Delivered Care Profile
  - **Declaration Dataset Code** Code referring to the (internal) table which contains the different products the grouper can deduce.
  - **Care Activity Code** The ZA code referring to external "care activities" table
  - **Amount** Amount of times the activity is performed
  - **Selling Price Hospital** For activities that can be reimbursed separately: code referring to external "tariffs" table which contains information on reimbursement amounts for that activity
  - **Selling Price Honorarium** For activities that can be reimbursed separately: code referring to external "tariffs" table which contains information on reimbursement amounts for that activity
  - **Declaration Code** For activities that can be reimbursed separately: code referring to external "tariffs" table which contains information on reimbursement amounts for that activity

8.3 Limitations of the DIS Data for efficiency measurements

A certain amount of limitations of the DIS data with regard to efficiency measurements can be derived from the DIS data structure and the interviews. These limitations are described in this section, after which the next section describes the possibilities of the data.

8.3.1 No traditional input factors, but throughput

Hospitals perform a series of activities, which results in a care product. In this sense, the activities could be seen as input factors for efficiency research. However, the activities cannot be compared to the input factors (economic resources) used in ”traditional” (scientific) research on hospital efficiency. Those researches use (costs spent on) staff, capital and materials, possibly further decomposed. With the DIS data, this decomposition into the *economic resources* is missing. The following quotes emphasize this:

> *”I do not see anything at all that I could use which has something to do with inputs”* [e16:sci].
> *”I see the DIS data mainly as an output measurement <...> Maybe activities as input.”* [e17:sci]

The activities (of a DBC) in that sense can be seen as ”throughput”. An unknown combination (black box) of economic resources is used to perform an activity, where a combination of these activities generates output (either measured by the total reimbursement or the total production).

8.3.2 Cost prices per activity are not directly in DIS, and potentially irreliable

Hospitals do calculate the cost prices for one activity (verrichting). These cost prices are derived from the total hospital costs (Dutch: *toerekeningsmodel*), both direct and indirect costs, which are decomposed with some formula (Dutch: *verdeelsleutel*) to individual activities [e2:ha, e7:bi]. There are however two large issues, related to the accessibility and reliability of these prices:

1. DIS does not require these activity cost prices, and there are no hospitals that deliver these [e8:bi].
2. Even if the (hospital specific) cost prices are available, it cannot be certain whether the cost prices of an activity are a good representation of the true cost prices [e7:bi]. Cost prices for a single activity are difficult to approximate from the different total costs made by hospitals and the many involved stakeholders [e16:sci], as shown by different studies on this subject (e.g., [3]).

National (average) Activity Cost Prices cannot be used either Although activity cost prices are not in the DIS dataset, these are collected from a certain amount (22) of hospitals by DBC Maintenance, the organization responsible for the creation of new products and the reimbursement rates. DBC Maintenance does not distribute these hospital specific prices (this would not be in line with the Dutch managed competition), but it does calculate average cost prices from these hospitals. These prices could, with permission, be available for research [e8:bi].

These national cost prices, however, are not very useful for ”our” efficiency calculations, as the interesting part of the calculations are the performances of a specific hospitals. Therefore, average cost prices of all hospitals are obviously not a good input measure. Theoretically however, it could be used to provide a certain weighting to an activity, so different compositions of activities could be compared

(approximating allocative efficiency). There is however one large drawback to this, which has been explained during the expert interviews with two of the heads of hospital administration [e2:ha, e5:ha]:

The cost prices are based on a single activity. Recall that not all activities are relevant for the deduction of a care product (both in the DBC system and with the DOT grouper). Section 7.7.2 already describes that hospitals often choose not to register such activities (e.g., lab tests). In those cases, the hospitals count the costs related to this (irrelevant for reimbursement) unregistered activity on top of the costs of the activity that is registered (e.g., consultation). In this case, one activity will look more expensive, but has actually the same costs as it involves multiple hidden activities [a comparable example of this is provided by two of the experts: e2:ha (page 29 of the interviews transcript) and e5:ha (page 76)].

> *"**You should** **never** **separate the costprice per activity from the corresponding profile.** However, this is done in DIS: they take the average national profile of twenty hospitals and couple that with an average cost price per activity of twenty other hospitals"* [e2:ha].

> *"You should actually require in the NZA model for cost prices that* all *activities in the DBC Maintenance activity table determine the costs. Instead of not registering half of the activities and attaching their costs to the other half. Which is what happens."* [e5:ha].

With this lack of a good (useful) cost decomposition, or costs or other information at all to provide weightings to activities, it should be clear that allocative efficiency measurements are not possible with the DIS data. At least not in this form and without providing more attention to the aforementioned issues.

8.3.3 Limited reimbursement (revenues) information

Besides limited information on the costs and decomposition in economic resources, reimbursement information is not available from the DIS data either.

An exception to this is a limited amount products for which the prices are non-negotiable: the products in the A-segment (B-segment is negotiable). For these sub trajectories, in the DBC data is a reference to the "tariff table" which states the prescribed (maximum) reimbursement. Although these products usually make up for around 30% for the total hospital revenues[4], these are not really useful either as they refer to specialized healthcare (e.g., brain surgery). And the possibilities of efficiency measurements, (e.g., comparing treatment approaches) are especially on the more common treatments which are often applied. For the increasing number of DBCs from the B-segment, price negotiations between hospitals and insurers are confidential (for sake of the managed competition).

[4] See e.g. http://www.nza.nl/133167/133442/458247/CI_12_70c-Kostprijsmodel-medisch-specialistische-zorg.pdf

8.4 Possibilities of the DIS Data for efficiency measurements

8.4.1 Aspects to take into Account

From the interviews, some side notes to take into account when using the DIS data for both efficiency and safety analysis were identified:

Analyze DBC (sub-trajectory) Transcending Analyses should not be performed on individual sub-trajectories (DBCs) but on complete "care paths" (Dutch: *Zorgpaden*), a combination of multiple subtrajects part of the same treatment for one patient [e2:ha, e5:ha, e14:con]: *"Don't look at a product or DBC, but at the diagnosis and all products related to this"* [e5:ha].Reason for this is that the separation into sub-trajectories (DBCs) is purely due to the financial system, while the complete treatment is a combination of multiple of those sub-trajectories (in the DIS set already registered under one care path). It should be noted that chronic diseases are part of a care path which only ends at the patients decease and are thus less suited for these kind of analyses.

Take Casemix into account Not all patients are the same, but they can roughly be categorized. Important aspects for this are age (year of birth is present in the DIS data) and comorbidity. Regarding comorbidity, the following quote describes another current issue with the DIS dataset: *"that is a deficiency in the system: there is 1 diagnosis and the comorbidity of the patient is not in there. <...> The possibility was created for this but I do not think anyone ever delivered one side-diagnosis to the DIS"* [e3:ha]. Efficiency measurements in healthcare should not be separated from the type of patient and medical content [e3:ha].

Do not compare different types of hospitals Another important part of this casemix differentiation, is not to compare university hospitals to general hospitals: the patients in a university hospital are usually more complex [e1:ha, e2:ha, e3:ha, e7:bi, e8:bi, e13:hc].

Stakeholders Perceptions differ The vision on a certain indicator differs per stakeholder. Within hospitals, this is specifically the managers versus the medical specialists. It should be kept in mind that benchmarks should be context independent and thus based on facts as much as possible [e12:hc].

8.4.2 Comparing Care Profiles (Technical Efficiency)

Taking this aspects into account, despite of the limitations of the DIS data, useful efficiency measurements can be performed based on it. These cannot be performed on the level of allocative efficiency due to the lack of information on the decomposition of activities in different (economic) resources. However, technical efficiency measurements are possible. For this, the (amount of) activities performed for a specific DBC (or careproduct) are compared between hospitals [e2:ha, e5:ha, e16:sci, e17:sci]. Obviously, as stated in the previous sub-section, due to the DOT system and DBC closere rules, these analyses might have to be performed DBC transcending, on a carepath level.

> *"Then also the "law of big numbers" is going to count. If you are a comparable hospital and have your patients come three times to the policlinic instead of two for a same operation, then for each hospital that are hundreds of operations from which an average number is computed that will significantly deviate."* [e5:ha]

Besides taking averages (amounts of activities performed for one product), also other statistics such as variance and outliers are very interesting to analyze. This was also shown from the car industry: a higher variance, many deviations from the average, was an indicator for an inefficient process [e14:con].

Categorizing activities However, there is another issue when analyzing these different activities and their frequencies of occurrence in the different hospitals for similar treatments. This is that there are so many different activities, that analysis on this level will result in too fragmented results. For example, hospital A might perform activity x^1 twice and activity y^3 once, while hospital B might perform activity x^2 four times, while performing activity y^4 once. With no additional information on the activities, no comparative statements on the production or efficiency of carepath can be made. For example, if it is known that all activities x^i and y^i are very similar in terms of cost prizes and/or reimbursements, it could be concluded the second hospital is less efficient.

Therefore, for efficiency measurements based on the activities performed per carepath, these activities will first have to be grouped into a limited amount of comparable activities.

- A first opportunity would be to categorize the activities per specialism, and further into ”attention areas”. These attention areas which would be developed together with medical specialists to achieve areas consisting of comparable activities, also medically recognizable. This categorization is already made in certain hospitals to analyze productivity and efficiency internally [e7:bi].

- Another possibility is to use care profile classes (Dutch: *”zorgprofielklasses”*). These are a very limited set of care activity categories (10-20, summed in Appendix D). This categorization is independent of the specialism the activity is performed by (/for): the categories are like ”operating activity” or ”policlinical visit”. These care profile classes reduce even more activity information than the previous option, but for the purpose of efficiency measurements this is no issue as more information would only lead to unambiguity [e8:bi, e16:sci] (see also e.g., [86]).

Categorizing in care profile classes (Dutch: *”zorgprofielklasses”*) is the most preferred categorization option of the two identified above. Reason for this is that these care profile classes are developed by DBC Maintenance and these are already used by this organization to create ”average care profiles”. These are the average amount of each activity (categorized in care profile classes) performed for the different DBCs, based on all hospitals’ data: care profiles. These analyses are fairly similar to the one described above and could indeed serve as a basis for efficiency measurements (to measure efficiency against) [e8:bi]. However, DBC Maintenance does not yet provide a separation between care profiles of academic and general hospitals, while academic hospitals generally receive more complex patients [e1:ha, e2:ha, e3:ha, e4:ha, e5:ha, e8:bi].

Comparing against a production function Another issue is that the care profiles provided by DBC Maintenance only allow for measurement against the average profile. As described in Section 4.2, efficiency measurements generally measure against a (production) frontier: the ”best of class”. The desired situation is therefore to measure against ”most optimal” care profiles instead of the average calculated by DBC Maintenance [e8:bi]. An even more desired situation would be not to base this frontier on historical (DBC) data, but draft it in collaboration with medical experts, in order to safeguard the quality of the provided care as well [e8:bi].

Further analyses of measurement results After performing efficiency measurements, further analysis for the interpretation of the results will be required. This would have to be performed by interviewing medical specialists from ”interesting cases” (e.g., specialisms with low efficiency score) [e5:ha]. The main topic for investigation would be why certain activities are relatively often or seldom performed. This deeper analysis is required, as it could be the case that some hospitals only perform an operation when absolutely necessary, while others might perform the same operation also preventive. Naturally, the first hospital would have only ”heavier” patients which would result in a lower efficiency score [e5:ha].

Another important aspect is that hospitals could not register (although performing) activities ”irrelevant” to the declaration. This phenomena is described in Section 7.7.2. For these hospitals, a higher efficiency score would be achieved, which is clearly undesired. Therefore, a validation on this population completeness would first have to be performed, a recommendation also described in Section 7.7.2. This possibility would also have to be taken into account when further analyzing efficiency scores.

8.5 Conclusion of the Efficiency Contextual Data Quality Analysis

This section has described the analysis of the expert interviews on the contextual data quality of DIS to evaluate efficiency. Of the nineteen performed interviews, six have been specifically on this data quality dimension although the other interviews often related to the usefulness of these data to evaluate efficiency as well.

Although some information and possibilities are missing, there are some very useful efficiency related evaluations possible with these data. With the DIS data, a cost-breakdown in traditional economic resources (materials, staff, capital) is not possible as only the activities are stored. Furthermore, cost prizes are either not available or not reliable enough for the envisioned analyses. Therefore, the commonly used allocative efficiency calculations in scientific research are not possible. However, there are some good opportunities of these data to evaluate technical efficiency, on the level of productivity of hospitals. Most notably, due to the registration of the performed activities, these data allow to compare (average) approaches of hospitals for a specific care product. For this, it will be useful to group the activities to care profile classes, which allows for a more uniform comparison. Important there, is not to compare university hospitals with general hospitals but to differentiate the three existing types of hospitals in the Netherlands: general hospitals, top-clinical hospitals and university hospitals. Another important aspect is not to separate the activity from the care path it was performed in as some hospitals interpret certain activity codes differently than others and could very well combine multiple activities in one code where other hospitals register these separately.

9 The Contextual Data Quality of DIS to Evaluate Patient Safety

The previous section provided an analysis of the contextual quality of the DIS data in relation to efficiency research. As contextual quality refers to the ”fitness for use” or ”usefulness” of the data in a specific context [57, 74, 85], it described the usefulness of the data to perform certain efficiency measurements. This section provides an analysis of the contextual quality of the DIS data in relation to patient safety research. Thus, to the usefulness of these data to perform analyses on patient safety levels in a hospital. This aims to answer the second research question of this thesis research: *”What is the contextual data quality of DIS to evaluate patient safety (underlying quality of care) in Dutch hospitals and how is this determined by intrinsic data quality of DIS?”*.

For this, an approach comparable to the analysis of context quality for efficiency research exists. First, incident identification methods are analyzed in Section 9.1. At this point, all data fields of the DIS data are potentially useful in the patient safety analyses, thus no section is provided describing the relevant data contents. After the description of the currently used patient safety measurements, the limitations of the DIS data to perform such measurements are described in Section 9.2. The possibilities are then described in Section 9.3.

As described in Section 4.3, two different approaches exist to analyze patient safety: prospective and retrospective [e12:hc]. The prospective approaches analyze the activities performed to minimize the chances for incidents, and thus is more a risk analysis. The retrospective approaches analyze incident reports after these have occurred. For these retrospective approaches, different levels of granulation exist for the analysis of the incidents. It can be high level, summarizing occurrences of incidents (and related events), or lower level, analyzing the individual incidents to discover causes and prevent future occurrences of similar incidents. This thesis research focuses mainly on retrospective analysis (see Section 9.2.2 for the reason behind this) and on what levels of granulation such analyses can be performed based on the DIS data.

9.1 Methods to Measure Patient Safety

As described in Section 4.3, patient safety refers to ”a situation in which the patient will not suffer, or has only a slight risk of suffering any damage caused by health care professionals who are not acting in accordance with the professional standards or by failures in the care system” [83]. This is very related to incidents: incidents are unintended happenings with a chance of damage. In case this damage does occur, the incidents are referred to as adverse events. This negatively influences the patient safety of a hospital. Therefore, the starting point of most patient safety researches is the analysis of incidents.

Before analysis of incidents is possible, these first have to be recognized. For this, hospitals are often dependent for this on voluntary incident reports made by employees. During the interviews, two approaches of recognizing incidents have been extracted, which allow for retrospective patient safety analyses. The first approach is used in one of the Dutch university hospitals to analyze the safety of the hospital internally. This approach is illustrative for many internal hospital projects, and is further described in Section 9.1.1. Secondly, there is the scientific approach which compares the safety of hospitals on a national level. This approach is described in Section 9.1.2. Besides analyzing incidents, another retrospective approach to analyze patient safety is by measuring mortality rates of hospitals. Furthermore, there currently is a large national project to analyze the hospital’s patient safety prospec-

tively: the ten safety themes. These last two approaches are shortly described in Section 9.1.3.

9.1.1 Internal Hospital Safety Projects

Currently all hospitals run a safety program which, as part of the program, allows for the reporting of incidents by internal staff. This section describes the incident reporting method of one of the Dutch university hospitals. This description is based on the explanations provided by one of the interviewees (e13:hc), who is employed to the concerned hospital as an internal quality manager.

In this hospital, the risk management system in use to report incidents and near misses has been implemented two years ago and is completely decentralized. This implies that *"when something goes wrong or almost wrong on a department, which can be medical, related to nursing, but also in the laboratory or X-ray, it is reported in that specific stand-alone system"* [e13:hc]. Each of these separate systems contains a risk matrix in which a grading for the reported incident is assigned, on a scale 1 to 4. All incidents are analyzed by decentralized commissions, after which it is their responsibility to perform correct actions towards the reporter and the incident itself. The more severe incidents (level 3 and 4) are also reported centrally. The central commission advises the decentralized commissions whether or not the incident is so severe it has to be regarded as a calamity and the calamity protocol has to be followed. The calamity protocol ensures these incidents are reported to the Dutch inspectorate (IGZ) and stakeholders such as head of department and the board of directors [e13:hc].

Safety Culture "veilig incident melden" For these internal safety management systems it is very important a specific culture of safe reporting exists for the employees with no risk of personal consequences. This has already been identified (and initiated) in literature, which is described in Section 4.3. Deriving from the interviews, it has also become an important precondition for the effectiveness of safety management systems as considered in practice [e12:hc, e13:hc, e11:dis, e18:sci].

> *"You want a specific culture in the hospitals to allow the employees to report incidents safely. If you ruin that culture <...> he might consider the next time not to report everything any more".* [e13:hc]

9.1.2 National Safety Benchmark Projects

The national safety benchmark projects such as the *monitor health related injury* [56] do not use the voluntary incident reports from hospital employees. For these researches, complete patient records (Dutch: *Dossiers*) are analyzed by nurses and experienced specialists.

In this approach, a sample of patient records is drawn from the hospital information systems. The nurses then apply a trigger tool on each record. The trigger tool includes sixteen triggers [56, page 34] which are all possible indicators of the occurrence of an incident. For example, one of the triggers is whether or not an unexpected re-admission took place. Appendix E provides a full overview of these triggers. The nurses do not identify incidents themselves, but with this triggertool the records with an increased risk for the occurrence of an incident in that treatment are identified.

For the next step, independent medical specialists analyze these "high risk" records. Three phases are performed to investigate damage. First it is determined whether or not there was damage. Second, whether this damage has resulted in decreased health to the patient. *"So it should have resulted in a*

longer hospitalization, physical injury, a wrongly amputated leg, that category" [e18:sci]. The third step is to determine whether or not the damage was caused by the care itself (e.g., not the disease itself).

This method results in statistics on the occurrence of incidents, adverse events and preventable adverse events in hospitals. The whole process is very labor-intensive. *"On average, it takes a nurse half an hour to analyze a record, and possibly another half an hour extra for the analysis by the specialist."* [e18:sci]

9.1.3 The National Mortality Rates and Ten Safety Themes

The Dutch inspectorate (IGZ) currently uses three main indicators for patient safety. Two of these are retrospective. The first retrospective approach is described in the previous sub-sections. For this, first incidents are identified, either by analyzing patient records (Section 9.1.2) or through an incident reporting system (Section 9.1.1). Then the incidents are further analyzed on severeness (for benchmarking) and causes (for internal improvement).

The second retrospective approach are the national (standardized) mortality rates. This analysis determines for each hospital the expected mortality rates, based on many factors such as academic setting and average patient's age. It then maps this to the mortality rate as occurred in that specific hospital. These rates are currently based on the LMR (Dutch national medical registration). However, as the LMR data was no longer mandatory for hospitals to deliver at the introduction of the DBC system, the results are not reliable for all hospitals [e13:hc].

The third set of indicators used by the IGZ are the ten different safety themes [49]. These are prospective and aimed to minimize the risks for incidents. Examples of the themes are "the early recognition of a vitally threatened patient" and "medication verification at admission and discharge".

Besides these indicators, other health care quality indicators exist, such as the 43 *Zichtbare Zorg* (institution for transparent healthcare) indicators. As patient safety is part of quality of care, these indicators to a certain extent also refer to safety issues. Usually, these are prospective indicators: describing the steps and actions taken to minimize the risks for incidents.

9.2 Limitations of the DIS data for Patient Safety Measurements

Before delving deeper into the possibilities of the DIS data for patient safety analyses, there are some important limitations which have been identified from the interviews and should be realized. These limitations reduce the feasibility of distracting incidents and other safety indicators from the data, and are further described in the following subsections (9.2.1 to 9.2.4).

9.2.1 Quality of the provided care not the focus of the dataset

One of the main limitations and the reason for certain more specific limitations is that the DIS data is a *financial* administration, resulting from a (financially focused) reimbursement system [e2:ha, e11:dis, e12:hc, e13:hc, e15:con]. Therefore, only aspects relevant to receiving reimbursements and for the maintenance of the system are stored and can be used for analyses. Table 18 provides some quotes stated during the expert interviews that confirm and further describe this.

Expert	Quotes Related to the Financial Focus
e15:con	*"Take into account that it has been developed from a financial point of view. DBC, and DOT, is a financial system, developed from the need for a new reimbursement system."*
e12:hc	*"It is a hospital administration goal. Which is, by definition, different than <eh> quality or safety as a goal. These systems are not developed to extract quality or safety indicators; but distinct systems that are not administratively related."*
e2:ha	*"A problem is that it is especially financially focused. So you are steering to thóse activities that mean something in the DBC game. So less on tangible quality indicators."*
e11:dis	*"There are no specific codes whether an incident has occurred. This is not relevant for the reimbursement."*
e11:dis	*"It <DIS> gives you a lot of information but I doubt whether it is relevant for safety management. Not as it is. <...> It is financial information, not directly useful."*
e13:hc	*"I do not think it is in there. At least not in this hospital, it is really a separate system. Everybody knows that here, incidents are reported in the risk management system, this is for the health care administration."*
e12:hc	*"A lot of this patient safety information, apart from volumes, can not be derived from the basic health administration, as they are stored in other separate systems. Partly because those systems, with regard to safety very consciously, are not freely accessible."*

Table 18: Interview Quotes describing the Limitations because of the Financial Focus

9.2.2 Some essential information on incidents is missing to derive patient safety

> *"If you simply think what is in that DIS, it is nothing else than a starting and ending date with a unique typing for the care path, with on the path several activities following an activity code. That is the essence of all that is in there."* [e2:ha]

The DIS data only contain performed activities on patients. As described in Section 4.3.1, not all incidents result in damage (for example an extended hospital admission or physical injury). Those incidents will thus not result in extra activities. This implies that only those activities resulting in damage (adverse events) could possibly be derived from the data. For example, *"You will probably not recognize in the DBC data whether a patient has filed a complaint. Those are typically aspects you get from the patient records."* [e18:sci]. As long as these incidents did not result in extra activities, these cannot be recognized from the DIS data.

> *"It is also possible the patient has fallen out of bed the second day. Or received the wrong nutrition. Or has called three times at night while the nurse did not show up... you will never be able to get those kind of incidents from this administration. It will require a more in-depth analysis in other administration, you will really have to perform analysis on patient records."* [e13:hc]

With regard to this, there is also a (population) completeness issue. Two of the interviewees have indicated that, even though an extra activity might have been required to remedy an incident, this might not be registered [e11:dis, e13:hc]: *"The patient is recovered, but the question is whether an extra activity is registered for that"* [e11:dis]; *"You will not find in these data when a leg has been amputed from the wrong patient: we will do that on own accounts instead of sending an invoice to the insurer"* [e13:hc]. An approach which is aimed to recognize incidents based on the performed activities described in the DIS data should take this possibility into account.

Processes to minimize risks of incidents are not part of activity codes Prospective risk analysis (i.e. measuring actions that are performed to minimize chances for incidents) is not possible with the DIS data [e12:hc]. This category of analysis is performed on more detailed information on the care processes and not specifically on the performed care activity. For example, safety markings can be applied to ensure instruments that need to be sterile are placed in the sterile airflow which exists in every operating room [e12:hc]. But this cannot be read in the administration of the performed operation, which is further limited by the ZA activity codes (see Sections 6.1.1 and 7.2.1).

As described in Section 9.1.3, the Ten Safety Themes provided by the Dutch Healthcare Inspectorate (IGZ) are typically prospective. Thus, the indicators contained in these themes cannot be measured with the DIS data [e12:hc, e13:hc].

9.2.3 Causes for activities are not stored

There is another large limitation to the DIS data. As stated, only the performed care activities are stored. These are registered with the ZA activity codes. Examples of these codes are 31250 which means "intracapsular cataract extraction (ICCE)" and 31268 which means "extracapsular cataract extraction (ECCE)" [22, DIS verrichtingen stamtabel v20120329]. More detail is not provided by the activity codings. The reason **why** the intracapsular or extracapsular approach has been chosen is not stored in the DIS data. However, this is relevant when the aim is to recognize incidents only based on the performed activities. An example used often is whether breast cancer is treated first with a breast-saving operation, after which still a breast-removing operation has to be performed [e18:sci]. This could indicate that the situation has not been interpreted sufficiently resulting in damage to the patient (extra operation, extended hospitalization). However: *"Yet another example. This is often what the women chose themselves"* [e13:hc]. Many women choose the breast-saving operation if there is only the slightest chance of recovery. The question is whether it is an incident when the recovery does in those cases not occurs as planned and an extra operation is needed.

This issue has been addressed by almost all interviewees when discussing the possibilities on patient safety, as shown in Table 19, which provides more examples. When limiting the safety analysis on incident discovery and (high level) retrospective analysis, this is still a large issue.

Expert	Quotes Related to the lack of causes
e15:con	*"You do not know the cause of say extra hospitalization. For example, are those caused by something irrelated such as an eighty year old choking in some food and getting in shock, or because of bedsores after the patient wasn't cleaned for one day."*
e11:dis	*"Do you then have a more expensive product because you have a more complicated patient, or because you have dropped him? <...> You will have to inspect what happened in the individual cases before you can draw conclusions "*
e18:sci	*"But you never see in the DBCs, and not in the patient record either, what the background was, what were the causes"*
e13:hc	*"In those cases you could ask yourself "what has happened here"? But then you would probably still need to analyze separate systems to determine what exactly happened"*

Table 19: Interview Quotes addressing the lack information on activity causes

9.2.4 Timeliness

An average timeliness of one year and one month (Section 7.4) is considered by most interviewees as "just too late" [e11:dis, e18:sci, e13:hc] or "fine for retrospective analyses" [e12:hc]. Half a year is considered ideal: enough time to be sure no later admissions were needed after (e.g.) an operation, but not too much time [e11:dis, e18:sci]. For discovering *trends and developments* in incident occurrences, the timeliness is not considered a large problem [e13:hc, e18:sci].

9.3 Possibilities of the DIS data for Patient Safety Measurements

9.3.1 Aspects to take into account when using these data

As described in Section 8.4.1 some side notes to take into account when using the DIS data for both safety and efficiency analysis were identified from the interviews. First of all, analyses should not be performed on individual sub-trajectories (DBCs) but on complete "care paths" (Dutch: *Zorgpaden*). Second, the casemix (type of patient) should be taken into account with as important consequence university and general hospitals should be analyzed and compared separately.

9.3.2 Discovering *possible* incidents in the Data using triggers

The approach described in Section 9.1.2 which uses a trigger tool (see Appendix E) to determine potential incidents in (records of) treatments, could also be used on the DIS data [e18:sci]. There are however a couple of issues related to this current trigger tool: it is specifically designed for the analysis of patient records (the core problem); it triggers in many cases on something that has happened unintentionally which is not stored in the DIS data (causes); and some triggers are based on activities or events for which no ZA activity code exists, which are thus not in the DIS dataset. Furthermore, the triggertool is only the first step in analyzing records on the occurrence of an incident; the actual incident recognition is in the next step performed by expert analysis: *"If you really want to see whether damage*

occurred, you will have to look into the specific record" [e18:sci]. A quote from another interviewee, when confronted with this idea, further confirms this:

> *"You could indicate which care paths or DBCs have an increased chance for incidents based on these data, but that is too short-sighted. It could be caused by so many things, you would have to look deeper into the data and these data do not provide that possibility"* [e13:hc].

After the triggering of a possible incident occurrence, no further investigation of the care path is possible. There is no more data in the set than already analyzed and combining with more detailed set is not possible. This is impossible due to privacy restrictions: a pseudonymized patients name would be useless if it was possible to trace the treatment back to the complete care record.

However, *"it might not imply that something has certainly gone wrong, but it does stand out <...> This is also what the IGZ uses, they say 'that stands out, show us your detailed data'. Then they start an inspection"* [e18:sci]. Most of the current safety indicators do not directly require information on (frequencies of) occurred incidents but merely use these kind of indicators.

As described before, the current triggertool is in this form not directly applicable to the DIS set. Other triggers should be developed for this. It is recommended to focus on those incidents that happen a lot. An example was provided as an abdominal surgery shortly after the removal of an appendix (usually an indicator some infection occurred in the abdomen). This approach is also used by the Dutch Inspectorate [e18:sci].

9.3.3 Datamining: discovering deviations from "standard" carepaths

The interviewees were predominantly negative on the possibilities of "reading" incidents directly from the DIS data. However, more faith was in combining the information provided for one carepath and, using data mining techniques, searching for notable deviations from other carepaths. These carepaths should be comparable, e.g., taking the type of hospital and patient's age into account [e2:ha, e11:dis, e12:hc, e13:hc, e18:sci].

> *"Then you are getting close to quality issues. Differences in treatment. But you are not going to query those in a reasoned manner. With that I mean you are not going to look at one specific field to find what you are looking for. No you consider the complete data and let it review automatically. Then I give it a chance."* [e2:ha]

Obviously, the causes of certain "deviations from the standard" can still not be derived from the data. Not every significant deviation from an average carepath (or protocol) implies the occurrence of an incident, and medical specialists should remain the possibility to depart from the standard path [e13:hc]. However, in line with most current safety measurements, the deviations do provide an incidation that the patient safety could have been at risk (i.e., an incident could have occurred).

Another issue is how to approximate and interpret a deviation from the standard [e15:con]. In a recent Master's Thesis describing the first steps of standardizing care paths, this has been addressed by using and analyzing different *distance measures* to recognize (and cluster) similar treatments [86].

Some examples of possible results provided by the interviewees are: unexpected longer hospitalization [e18:sci]; the timespan between a first consult and an operation [e2:ha, e5:ha], and erroneous hospital discharges [e18:sci].

9.4 Conclusion of the Patient Safety Contextual Data Quality Analysis

This section has described the analysis of the expert interviews on the contextual data quality of DIS to evaluate patient safety. Of the nineteen performed interviews, six have been specifically on this data quality dimension although some other interviewees often referred to the usefulness of these data to evaluate patient safety as well.

As with the usefulness of the DIS data to evaluate efficiency, some interesting opportunities of the data to evaluate patient safety exist. However, there are also some restrictions in the possibilities which are mainly related to the strong financial focus of the dataset. Most notably, prospective analysis of patient safety is not possible (i.e., what precautions are made to minimize the chances for incidents and thus improve patient safety).

Most opportunities of the data are on retrospective analyses (i.e., analyzing what incidents took place). This information on incidents is not directly available in the DIS data, especially as the reason for performing a certain activity is not included in the data (e.g., was the activity performed as the consequence of an incident or was this already planned beforehand).

Data mining on the DIS data could be used to recognized unexpected patterns for certain (oftern performed) care paths. From this, a high chance of the occurrence of an incident can be identified on the level of care paths. Such numbers could be used by the Dutch healthcare inspectorate (IGZ) as a reason to observe a hospital more closely for a specific amount of time. As with efficiency evaluations, casemix and type of hospital will have to be taken into account. Furthermore, the purely financial separation into sub trajectories has to be removed by looking, DBC transcending, to the whole care path.

Part III

Conclusions

10 Conclusion

This research focuses on the data quality of a large Dutch national healthcare database: the DBC Information System (DIS). This database contains data relevant to the Dutch DBC reimbursement system. The main research question is formulated as *"What is the intrinsic and contextual data quality of the DBC Information System within the Dutch hospital care, and how are these determined and interrelated?"*. Here, intrinsic and contextual quality are two important quality dimensions often used in literature.

To answer the research question, three sub-questions have been drafted. The first sub-question is to determine the intrinsic data quality: the answer to this question is provided below, in Section 10.1. The second and third sub questions relate to the contextual data quality. Many contexts are possible; for this research the evaluation of efficiency and patient safety have been elected due to their practical and scientific relevance. The answer to the second sub-question, on the contextual quality with regard to evaluating efficiency, is provided in Section 10.2. The answer to the third sub-question is provided in Section 10.3, which focuses on the contextual data quality with regard to evaluating patient safety. Section 10.4 concludes with the final answer to the research question.

10.1 Intrinsic Data Quality from a process perspective (sQ1)

The first sub-question was formulated as follows: *"What is the intrinsic data quality of DIS and how is this determined by organizational and technical issues?"*. Instead of a quantitative approach on the data to determine the intrinsic quality, a qualitative process-based approach based on expert interviews was used. This process-based approach aimed to identify (organizational and technical) process issues eventually influencing the data quality. Besides providing a good overview of the intrinsic quality, this approach also allows for improving these process issues. The three most important issues found, and their impacts on the DIS data quality, are described below:

Only mandatory data relevant to declarations are stored The DBC (and DOT) system is a financial system which implies that also the DIS dataset is financially focused. Because of this financial focus, especially with the introduction of the DOT system, it can be expected hospitals register more complete. After all, what is not registered is not reimbursed (activity based costing). However, this (population and column) completeness is only relevant to those activities and data fields that influence the final declaration, or that are required by the DIS set or grouper. Culture of medical staff towards administration also plays a role here: what is not mandatory to register, will rarely be done.

This implies that it has to be taken into account that many activities irrelevant to the declaration (e.g., lab activities) and data fields that are not mandatory (e.g., secondary diagnoses or ICD-10 codes) are likely to be missing in the DIS data.

Only Declared Products are in DIS This issue is also related to the aspect that the DIS dataset is the result of a financial system and has a financial focus. The care products that are sent to the DIS are only those that have been declared. Before declaration is possible, the sub trajectory (DBC) has to be closed and validated, and provided with a declaration code by the grouper. Furthermore, price arrangements with the health insurers are required.

All these aspects result in a timeliness (age) of more than one year, before a certain treatment is in the DIS database. Furthermore, in cases where a combination of diagnoses and treatment is not recognized by the grouper and cannot be declared, it is not stored in DIS (population completeness issue) or changed by the hospital to another product than actually performed for which a reimbursement is then requested (semantic accuracy issue).

Issues with the Aggregation of Performed Activities to Activity Codes In the DIS dataset, the activities are coded with ZA (*Zorg Activiteiten*) codes. The issue with this is especially related to university hospitals, which register their activities with the more detailed CBV codes.

The management of the CBV codes and the aggregation to (far less) ZA codes is problematic. The ZA codes are often insufficient to represent the medical complexity of the activity, which could result in decreased semantic accuracy of the data. Furthermore, due to the innovative setting of university hospitals, often a coupling between ZA and (newly created) CBV codes is not present yet. This could result either in decreased population completeness when the activity is in those cases not stored in the DIS set, or decreased semantic accuracy when another comparable ZA code is chosen for reimbursement purposes.

The described problems are less present in general hospitals as these perform more ”standard” treatments and (often) register with CTG codes (comparable to ZA codes), where university hospitals perform more innovative activities and treatments on more complicated patients.

10.2 Contextual Data Quality: Usefulness to determine Efficiency (sQ2)

The second sub-question was formulated as follows: *”What is the contextual data quality of DIS to evaluate efficiency in Dutch hospitals and how is this determined by intrinsic data quality of DIS?”*. For this research, this contextual quality is interpreted as ”usefulness”. Therefore, to answer this sub-question, the usefulness of the DIS data to perform efficiency measurements is assessed. With this regard, some limitations have been identified. However, although some information and possibilities are missing, very relevant efficiency-related assessments are possible with the DIS data.

A first limitation is that a costs breakdown into traditional economic resources (staff, capital, materials) is not possible with the data as only the activities are stored. Furthermore, cost prizes are either not available or not reliable enough for the envisioned analyses. Therefore, traditional input measures are not available from the data and the commonly used efficiency calculations in scientific research are not possible. Without information on the costs (or reimbursements) of DBCs and activities available, all allocative efficiency measurement (i.e., optimal combination of used resources) can *not* be performed with these data.

Measuring Technical Efficiency: Comparing Care Profiles However, the data are suited to perform technical efficiency measurements (i.e., minimizing input) with DIS. This is especially the case on a detailed level, comparing how different hospitals (or specialists) perform certain treatments. These analyses are comparable to efficiency measurements currently performed internal to hospitals, aimed to provide management information. An issue that remains is the high amount of different care activity (ZA) codes (although often still not detailed enough for medical analyses). DBC maintenance has already

provided a mapping to *care profile classes* (Dutch: *Zorgprofielklasse, ZPK*). With this, 24 different classes of care activities are identified, such as daycare or surgical activities.

Based on care profile classes, the amount of different types of activities hospitals on average perform for a treatment could be compared to the hospital that perform this most efficient. Important is not to compare different types of hospitals (e.g., university versus general hospital) and, where possible, to take casemix (i.e., type of patient) into account.

10.3 Contextual Data Quality: Usefulness to determine Patient Safety (sQ3)

The third sub-question was formulated as follows: *"What is the contextual data quality of DIS to evaluate patient safety (underlying quality of care) in Dutch hospitals and how is this determined by intrinsic data quality of DIS?"*. As with the second sub-question on efficiency, for this the "usefulness" of the DIS data to perform analyses on patient safety is assessed.

Different approaches of patient safety analyses exist: prospective (analysis of risks for incidents) and retrospective (analysis of incidents). Prospective safety indicators, such as the IGZ ten safety themes, cannot be derived from the DIS data. Reason for this is that more detailed information on processes and protocols is needed for this category of indicators. The DIS data only contain a code describing which activity was performed and not the detailed information on the processes or protocols followed behind the activity.

The retrospective approach consists of first identifying incidents and subsequently analyzing (root causes of) these incidents. Due to the financial focus of the DIS data, the latter phase of root-cause-analyses is not possible as detailed information from the patient record is needed for this.

Using Data Mining to Discover Incidents However, there are possibilities of recognizing (potential) presence of incidents. Not through conventional (and labor-intensive) methods (e.g., manual patient record analysis [56]), as causes behind activities are not stored in this set. The possibilities are in applying data mining techniques to the data to identify unexpected deviations in the treatment path. For this, aspects such as casemix and type of hospital have to be taken into account, and the purely financial separation into sub trajectories has to be removed by looking, DBC transcending, to the whole care path.

Comparable to most other current safety indicators, recovered potential incidents from these analyses might not directly prove lack of safety but merely provide an incentive of the Dutch health inspectorate (IGZ) to perform further inspections on the specific department. This will improve the efficiency and effectiveness of these researches as they are specifically guided to high-risk areas.

10.4 Answer to the Research Question

The main research question of this master thesis was formulated as follows:

> *"What is the intrinsic and contextual data quality of the DBC Information System within the Dutch hospital care, and how are these determined and interrelated?"*

To evaluate the intrinsic data quality of the DIS, this research has used a process-based approach. Some organizational and technical issues negatively influencing this intrinsic quality have been identified. Most notably, due to the financial focus of the DIS data all content irrelevant to the declarations are likely not to be in the DIS (population completeness). Furthermore, the data is only submitted after the whole declaration cycle has been performed, resulting in a timeliness (age) of over a year. Another large issue is that the codes used in the DIS data to represent activities (ZA codes) are very high-level. This results in issues on semantic accuracy due to maintenance of the activities and aggregation within the system.

With regard to registration of the data, most fields are restricted in terms of possible values and often refer to external tables. This greatly improves some dimensions of data quality, such as syntactic accuracy and column completeness. However, due to these restrictions in possible values and mapping problems such as described above (aggregation to ZA activities), this results in some serious quality issues with relation to especially semantic accuracy and population completeness. Furthermore, this research shows that, respecting time-limitations and priorities of the medical specialists, the data are entered in the system by the most appropriate employees. However, the time before this registration is performed can easily take up to a week (currency), where it would ideally be performed within one day.

Due to the qualitative approach of the research (opposed to a quantitative approach), no specific and measurable statements on the intrinsic DIS data quality can be made. However, based on the organizational and technical issues identified it can be concluded that improvements are certainly possible. Especially for university hospitals, the DIS data is not always accurate and complete in reflecting the delivered care. This is mainly due to the innovative and unique nature of many treatments performed. However, especially for the more standard treatments often performed (e.g., knee or hip surgery), the intrinsic data quality of DIS will be of such good quality that reliable conclusions to evaluate hospitals can be drawn based on the analysis of these data.

This reliability of analyses has for this research been focused on and scoped to efficiency and patient safety analyses. Contextual quality here is assessed as the ”fit for use” or ”usefulness” of the data to perform such analyses. For patient safety research, not all types of analyses are possible with the data: possibilities are mainly on more advance data mining techniques to discover incident occurrences in the data. These incidents correlate with patient safety. For efficiency research, there are some great opportunities with the DIS data. These are especially on technical efficiency measurements: comparing the care profiles different hospitals perform for similar treatments. There are however also some limitations with regard to allocative efficiency variants: a cost decomposition into economic resources (e.g., staff, capital) is not possible and information on cost prices and reimbursements for specific activities or products are not available or unreliable.

11 Recommendations

Based on the expert interviews and conclusions of this research, there are some recommendations to improve the intrinsic and contextual data quality of the DBC Information System (DIS). These regard getting underivable products in scope of the organization governing the data (also referred to as DIS); performing more validations on the completeness of the data; uploading also the CBV codes which were initially used to register the performed activities; and providing a closed feedback loop between DIS and the hospitals. The following sub sections explain these recommendations in more detail.

11.1 Underivable treatments should be in scope of DIS

Those activities and sub trajectories that could not be deduced by the grouper into a valid product (Dutch: *grouper uitval*), should be submitted to the DIS dataset as well. At this moment, the DIS set only contains declared products while (especially university) hospitals face the problem that many of their more innovative and complex treatments cannot be derived into a valid, declarable product. On the other hand, DBC maintenance mainly bases the creation of new, and maintenance of old, activities and products on the DIS dataset. They do also receive signals from hospitals on specific treatments that cannot be derived to a care product, but this is incidentally and only for real urgent issues.

Consequently, as long as the underivable treatments are not accepted by the DIS, this vicious circle will subsist and those treatments could also in the future not be derived. Obviously, hospitals would still have to analyze the underivable treatments to make sure no erroneous registrations have been performed.

As described earlier, many hospitals currently change activity codes or DBC composition when confronted with an underivable product in order to receive a declarable product from the grouper. It has to be taken into account that it is not likely this will change when also accepting underivable products. Ultimately, the priority of hospitals is more with receiving a reimbursement than with providing perfectly accurate data to the DIS database. Therefore, it might be required to accept both the original underivable combination of diagnosis and activities, and the later derived DOT care product.

11.2 Perform ”negative” validation on data completeness

Current DIS validations are mainly ”positive”: based on aspects that are in the data. Also ”negative” validations should be added to validate whether activities that would be expected are *structurally* not submitted. Hospitals do have the possibility not to submit certain information irrelevant to the declaration to the DIS. This cannot be detected if only incidentally performed. But hospitals can program their information system as such, that these kinds of activities are never submitted. In those cases this extra validation would aid in increasing the quality of the analysis results. For example, a hospital could seem to perform all DBCs very efficient, but if lab treatments are structurally absent, it should be compared to hospitals that do have lab treatments.

Another possibility is to develop certain ”expected profiles” for specific treatments, in collaboration with medical specialists. With this, minimum expected types and frequencies of certain categories of activities (i.e., care profile classes, ZPKs) could be developed. If hospitals do not meet these expectations, further analysis could be performed whether this is a structural data submission problem or not.

11.3 Including both CBV and ZA activity codes (university hospitals)

University hospitals (UMCs) currently register their activities with the more detailed CBV activity codes instead the ZA activities used for DIS and declaration purposes. As described in this thesis, many problems are currently experienced when aggregating these CBV codes to ZA codes. Therefore, the ZA codes as delivered by the university hospitals have a high risk of being semantically inaccurate. However, the DIS is unable to recognize these wrong aggregations, as the input information is no longer available.

It is therefore recommended, not only to send the ZA codes, but also the input on which these ZA codes was based: for the university hospitals the CBV codes. Especially if DIS accepts underivable products as well, DBC maintenance could then recognize unknown CBV codes and unknown combinations to ZA codes. Subsequently, it could examine whether the aggregation of the CBV to the ZA code is correct. Possibly, other aggregations or new ZA codes could be developed more efficiently with this new information. Furthermore, by including also the more detailed CBV activity codes besides the currently included ZA codes, a higher level of medically relevant detail would be derivable from the data. This will improve the possibilities of performing analyses on the data, especially with regard to patient safety.

11.4 Providing a closed feedback loop between DIS and hospitals

There is currently no closed feedback loop between DIS and the providing hospitals: most hospitals do not know how their own treatments appear in the DIS dataset. This thesis research has been performed qualitatively on stakeholders' perceptions of organizational issues influencing the quality of the DIS data. If a quantitative data quality analysis would be performed, reference material should be present. Without this closed feedback loop, this quantitative analysis could not be performed.

12 Discussion

Concluding this research, first a reflection on the findings will be provided by describing the practical implications in Section 12.1. This first reflecting sub-section will close with the limitations of the results. The subsequent sub-section then aims to explain some of these limitations by providing a reflection on the research method (Section 12.3). Future research opportunities are then discussed in Section 12.4.

12.1 Practical Implications

For this thesis research, a process-based quality analysis of the DIS data has been performed, and the usefulness and possibilities of these data for efficiency and patient safety research are examined. The results of this process are summarized in Section 10. However, there are also some practical implications behind these results.

First of all, some organizational and technical issues have been identified that influence the DIS data quality. Understanding these issues is important when interpreting analyses based on these data. One of the conclusions made based on the interviews is that some of the current analyses are in that sense not in line with the possibilities of the data. A good example is how the cost-prices are currently calculated. These calculations are performed on activity level, where these activities are disconnected from the corresponding carepath and analyzed separately. However, due to the way the activities are registered and how these should often be interpreted differently depending on the related carepath, this separation should not have been made. Therefore, different approaches on calculating cost-prizes (or categorizing activities) should be made for certain carepaths.

Another conclusion made based on the interviews is that currently a culture exists in hospitals to only register the necessary and required data. This is understandable, as the Dutch healthcare is in a situation where many cost-cuttings are made and where specialists have to prioritize their tasks to perform as effective and efficient as possible. However, with the current IS/IT possibilities, this should not be such an issue. In many hospitals, separate systems are in place for the internal administration of patient records, and for the financial registrations used for declaration. For each of these systems, a largely separate registration process is in place, and there is little communication between the different systems. The latter is also one of the main subjects of the most recent report of the Dutch health inspectorate [50]. If the different (hospital) information systems would communicate better or could even be integrated, much registration overhead would be released from the medical employees. Furthermore, the first (detailed) patient record administration involves all data relevant to the patients' treatment. With well integrated systems, no data irrelevant to the declarations would be missing from the DIS data because of the medical relevancy.

Furthermore, this research explored the possibilities to evaluate both efficiency and patient safety on the same dataset. It was found that many different approaches exist to measure or analyze these concepts, although these approaches are all separate and not related to each other. Efficiency research hardly takes into account the aspects of quality of care or patient safety, and vice versa. With the DIS data, some of the efficiency and patient safety measurement approaches are promising, although these are unrelated and due to the financial focus of the system, the efficiency measurements seem to be most potential. A desired future situation would be not to have a solely financial focus and be able to analyze both concepts based on the same data: to what extent can efficiency improve without jeopardizing

patient safety, and how can patient safety be achieved in a way as efficient as possible. For this, the financial and medical ”worlds” will have to be combined. However, the results of this research mainly show how separated these two concepts still are in practice.

12.2 Scientific Contributions

Besides the practical implications of this research described in the previous subsection, this thesis research provides several scientific contributions.

First of all, a data quality measurement was constructed for (this) healthcare related research. This had not been performed in Dutch healthcare literature yet, and frameworks in this context are also scarce in international literature. The framework was directly applied and tested in practice and after some slight adoptions (described in section 12.3.3) presented in this thesis document. This final version of the framework has proven very useful as it combines two important aspects of two older frameworks. First of all, a distinction between intrinsic and contextual data quality is made. Secondly, three quality dimensions often used (accuracy, completeness and timeliness) are divided into sub-dimension. The latter solves semantic issues on the definitions of these concepts often experienced in other literature (see e.g., [57]). During this research, this (final) framework has proven to be a clear and complete framework to guide the research on data quality, with a side note that two other dimensions of one of the underlying frameworks [85], representational and accessibility data quality, were not in scope of this research.

Furthermore, this research indicates that these data are useful to perform efficiency research on operational activities of the Dutch hospitals. Although some adoptions to the data have to be performed, especially on level of granularity, these can be done based on the data itself. Furthermore, it provides some more specific directions in which these data can be used for such research. This thesis would therefore serve well as a basis for scientific research on efficiency, based on these promising data.

This thesis document further describes the usefulness of these data for patient safety research, where this patient safety is an important factor determining quality of care. Some current limitations of these data were discovered due to the intrinsic quality of the DIS data and financial focus of the system. However, this thesis provides a thorough data and business understanding, a fundamental for further scientific research on extracting patient safety indicators from these data with data mining methods. This is further described as a future research opportunity in Section 12.4.

12.3 Reflection on the Research Method

The answers to the research (sub) questions are based on the results from expert interviews. A total of eighteen interviews (1 to 2.5 hours) were performed with various stakeholders related to this research. An interview protocol was drafted to guide the interview process, and a quality framework was composed to structure the interviews and analysis. The interviews have been transcribed resulting in a dataset of nearly 200 pages of qualitative data. For the analysis, a method similar to the *grounded theory* approach was used on this dataset. In Section 2.4, a description of the safeguards for Yin's validity threats [92] is already provided. This section provides a further, retrospective, view on the used research method.

12.3.1 Limitations of Qualitative Approach

Lack of quantitative measurements is an important limitation of this research, although this was at this state not possible. Some important organizational and technical issues were uncovered with the qualitative process-based approach from the expert interviews. However, these statements cannot further be supported by quantitative measurements, for example on completeness of the data.

Especially for the contextual data quality on efficiency and patient safety, possession of the DIS data itself would have been a great addition to this research. From the interviews, the perceptions of the interviewees on the data were uncovered and their views on the usefulness of the data for these analyses. However, the latter was mainly described from the point of view of the limitations of the data. Therefore, the overall image of the possibilities of these data, based on the interviews, is quite negative. Furthermore, the descriptions of possible analyses of the DIS data are very high level, as only ”basic ideas” of possibilities of the data were mentioned by the interviewees. In order to get more tangible results (analysis descriptions), access to the data itself is required. However, the possibilities of the DIS data specifically with relation to efficiency measurements are further explained in a thesis research performed parallel to this one [86]. This focuses on optimizing the carepaths (profiles) and uses a sample dataset. Although the quality of this dataset could not be compared to the DIS data, it did provide a general insight into the more concrete contents of the data.

12.3.2 Expert Sampling

An important aspect when sampling the experts for the interviews was to create a diversity of the interviewees. This was expected to result in a diversity in interview results, based on different stakeholder groups. However, when further developing the interviews, not all experts were found to be as suitable to be interviewed for all three of the thesis topics. Consequently, most interviews focused on one or two of these topics (intrinsic quality; efficiency; safety). With this, most stakeholder diversity was disappeared as well. To assess the intrinsic data quality, mainly hospital administrative managers were interviewed, while for efficiency and safety mainly scientists and either internal efficiency (BI) or quality consultants were interviewed.

Another issue is that, especially on the intrinsic data quality, only experts very ”close” to these data were interviewed. These experts were in almost all cases the heads of the hospital administration department of large hospitals and therefore very knowledgeable on especially the issues in the DIS data and related registration processes. On the one hand, this provides much confidence that all or at least most of the issues are described in this thesis. On the other hand, this might create a too negative impression of the data quality, as focus was only on the issues.

12.3.3 Used Data Quality Framework

To structure the interviews, a data quality framework was created based on an influential framework from literature and a book on data quality already combining many influential date quality publications. During the interviews, this framework has proven a very useful tool to structure the sessions. A disadvantage was that some overlap between the different dimensions existed, mainly with the believability, objectivity and reputation dimensions. The definitions, although adopted from literature, did not always provide a clear view on the boundaries of the different dimensions. This phenomena of overlapping dimensions is

also recognized in literature. This observation has resulted in a slight adoption of the framework halfway through the interviews where these three were combined into a ”trustability” dimension (also supported by one of the literature sources).

12.3.4 Differences DBC and DOT

Another issue discovered halfway through the interviews was the (expected) impact of the DOT system on the data quality. The differences between the DBC and DBC On the way to Transparency (DOT) system were early in the research identified and described in this thesis. However, the impact of the different systems was only identified during the expert interviews. This was too late to make a clear choice between analysis of the DIS data generated during the DBC system or those generated from the DOT system. It would have been more appropriate to clearly perform this choice, as the approaches to the data analysis differ for the two systems. For DBC, there is already for years of data in the DIS. For DOT, which was only introduced recently, except some test files no data is uploaded yet. Therefore, an analysis of the DOT data quality would be more hypothetically.

During the interviews, some quality dimensions were identified which were influenced mostly by the DOT system, compared to the DBC system: semantic accuracy, population completeness and timeliness. Especially during the later interviews, these dimensions were then specifically assessed both for the DBC and for the DOT system. For the analysis of the interview results, differences in the DBC and DOT system were reported where found. With this, at least a clear understanding of the differences of these systems and their impact on the data quality should be created.

12.3.5 Validity Issues

Yin [92] describes four validity issues relevant to empirical research. These criteria are internal validity, external validity, construct validity and empirical reliability. Section 2.4 already describes how the research method used for this thesis research safeguards these four validity issues. As described in that section, construct validity was safeguarded during this research by providing a thorough explanation of the concepts used in this research and their definitions (Section 4). The definitions are all based on scientific sources or official documents and where discrepancies exist, these have been described. Empirical reliability, the possibilities to repeat the research with the same results is always difficult to achieve with a qualitative research approach. However, with the interview protocol provided, the description of the expert sampling, and the used ”grounded theory” approach, this is safeguarded as well as possible. There are some extra observations of this research regarding internal and external validity not mentioned in Section 2.4: these are described in the following two paragraphs.

Internal Validity Internal validity refers to the possibility to establish causal relationships and distinguishing spurious relationships. To minimize this threat, triangulation was used. For this, experts were sampled from different stakeholder groups to allow different views on the data quality. Furthermore, within each stakeholder group, at least two experts were interviewed to allow comparison of the different quality interpretations.

During the interviews for each expert a specific focus was on only one (occasionally two) of the quality aspects examined in this research: intrinsic quality or contextual quality on either efficiency or safety. The other two aspects were also treated during the interviews, but only briefly.

It was noted that depending on the stakeholder group, and not on the expert, the focus of the interview was determined. For example, all employees related to hospital administration were interviewed for the intrinsic data quality, while internal hospital consultants were interviewed for their area of expertise: either efficiency or safety (depending on their role). Furthermore, there have been some exceptions in the interviews where not every stakeholder group was represented by two or more interviewees (see Section 5).

Although these issues could theoretically influence the internal validity of the research, 18 different experts were approached which still allows for good comparison (and triangulation) between their perceptions. These were usually in line with each other and although some experts had a more negative opinion on the entire DBC and DOT system than others, similar comments on the different data quality dimensions were provided. Where gradings of intrinsic quality dimensions differed, this was usually due to a somewhat other interpretation of the ”severeness” of certain issues and not on the organizational issues underlying these data quality issues.

External Validity The second validity issue is that of external validity, which is briefly described in the research method section (2.4). This refers to the generalizability of the results [92]. For this research, generalizability mainly refers to the extent in which the found DIS data quality reflects the quality of the data delivered by all Dutch hospitals instead only those directly examined. The DIS data quality in this sense mainly refers to the intrinsic quality. For this research aspect, six university hospitals and one general hospital were visited.

Consequently, as in multiple occasions specifically mentioned by the interviewees, most of the issues found are related to university hospitals. Some of these issues do not relate to general hospitals. For example, (most) general hospitals do not register activities with CBV codes, but with CTG codes. The ZA (care activity) codes used in DIS are directly based on these older CTG (Dutch: *College Tarieven Gezondheidszorg*) codes. Thus, the whole phase of aggregation from a more detailed activity code to a high-level activity code and accompanying issues is not really present in general hospitals. Whether a certain issue only refers to one or multiple of the hospitals types has been described during the analysis in appropriate phrases. During the interviews, many of the experts have stated that the university hospitals have more data quality issues. The main reasons for this are mentioned as the higher complexity of the provided care and the fact that in university hospitals the specialists are employed instead in a partnership and thus have less motivational incentives to register.

Concluding, not all the organizational and technical issues described refer to *all* Dutch hospitals, but especially to the university ones. However, when looking from a data point of view, I am confident all main issues influencing these data are described in this thesis.

12.4 Future Research

This thesis research has used a qualitative research method of expert interviews to assess the DIS data quality process-based. As mentioned earlier in this section, it would be a valuable addition to the insights gained on the data quality if the actual data was accessed and more quantitative tests would be performed. With these additional tests, expert statements could be validated and impact values of the different organizational and technical issues found influencing on the eventual data quality could be calculated. In many cases, the quantitative data quality measurement would be a research in itself, mainly due to lack of reference data.

This research has shown the potential of the DIS data to evaluate hospitals on efficiency and patient safety performance. Comparable to quantitative data quality researches, the actual development of such (efficiency and patient safety) measurement methods would be a research on itself. A master thesis project performed parallel to this research has performed an initial research on comparing care profiles, based on a sample set similar to what would be in DIS [86]. A combination of the two theses could serve as a basis for further hospital efficiency research based on the DIS data.

For patient safety, such related research has not been performed yet. As concluded in this thesis, there is potential of the data to recognize incidents or adverse events in the provided care by performing data mining. As patient safety is defined as a situation in which a patient will not suffer or only has a slight chance of suffering such incidents, this would result in a quantitative and objective measurement of patient safety. This research could serve as a basis for the necessary data and process knowledge needed for such data mining approaches. Considering this, this research could be regarded to cover the first two steps of the often applied *Cross Industry Standard Process for Data Mining* (Crisp-DM) [17,89]: business and data understanding. The process consists of six steps, which are visualized in Figure 9.

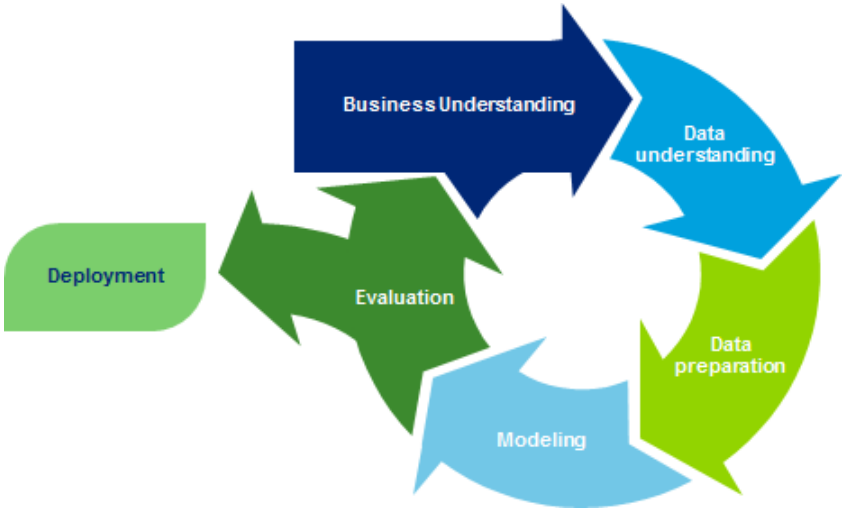

Figure 9: Diagram of the Crisp-DM Data Mining Method (Deloitte image based on [17,89])

Data understanding is created during this research by evaluating the intrinsic data quality process-based. Business understanding, understanding the business requirements for the analysis, is provided for the two fields evaluated with the contextual data quality (efficiency and patient safety). Accessing and preparing the data, and modeling different data mining approaches are left for future research, as is the evaluation of these methods and eventually eliciting and deploying the most suitable one.

References

[1] ADDICO, H. *Data and Information quality of XML databases in the healthcare domain*. PhD thesis, University of Sheffield, 2009.

[2] AERT, J. H. V. *Ziekenhuiskosten in econometrisch perspectief*. Stenfert Kroese, Leiden, 1977.

[3] ASSELMAN, F. *Kostprijzen in Ziekenhuizen*. Bohn Stafleu Van Loghum, 2008.

[4] AUERBACH, C. F., AND SILVERSTEIN, L. B. *Qualitative Data: An Introduction to Coding and Analyzing*. New York University Press, 2003.

[5] BARROW, M., AND WAGSTAFF, A. Efficiency measurement in the public sector: An appraisal. *Fiscal Studies 10*, 1 (1989), 72–79.

[6] BATINI, C., CAPPIELO, C., FRANCALANCI, C., AND MAURINO, A. Methodologies for data quality assessment and improvement. *ACM Computing Surveys 41*, 3 (2009), 16:1–16:52.

[7] BATINI, C., AND SCANNAPIECA, M. *Data Quality: Concepts, Methodologies and Techniques*. 2006.

[8] BLANK, J. Maak eind aan dot-operatie in ziekenhuis. *Het financieele Dagblad, Zorg en Markt* (July 5, 2010).

[9] BLANK, J., DUMAIJ, A., AND VAN HULST, B. Ziekenhuismiddelen in verband. een empirisch onderzoek naar productiviteit en doelmatigheid in de Nederlandse ziekenhuizen 2003-2009. Tech. rep., Centrum voor Innovatie en Publieke Secor. Efficientie Studies, Technische Universiteit Delft, 2011.

[10] BLANK, J. L. T., EGGINK, E., AND GRAAFF, A. I. D. Zuinig op zorg: Een empirisch onderzoek naar de productiestructuur van verpleeghuizen in nederland. *Zuinig op zorg September* (1996).

[11] BLANK, J. L. T., EGGINK, E., AND MERKIES, A. H. Q. M. Tussen bed en budget. *Den Haag: Sociaal en Cultureel Planbureau / Elsevier Bedrijfsinformatie* (1998).

[12] BLANK, J. L. T., AND MERKIES, A. H. Q. M. Empirical assessment of the economic behaviour of dutch general hospitals. *Health Economics 13*, 3 (2004), 265–280.

[13] BOEIJE, H., 'T HART, H., AND HOX, J. *Onderzoeksmethoden*. Boom Onderwijs, 2009.

[14] BORKENT, B., AND SCHEURWATER, R. Datadocument r21-dataset (version 0.4). Tech. rep., DBC-Onderhoud, April 2012.

[15] BRUIJNE, M. C. D., ZEGERS, M., HOONHOUT, L. H. F., AND WAGNER, C. Onbedoelde schade in nederlandse ziekenhuizen. dossieronderzoek van ziekenhuisopnames in 2004, 2004.

[16] CAMINITI, C., DIODATI, F., BACCHIERI, D., CARBOGNANI, P., DEL RIO, P., IEZZI, E., PALLI, D., RABOINI, I., VECCHIONE, E., AND CISBANI, L. Evaluation of a pilot surgical adverse event detection system for italian hospitals. *International journal for Quality in Health Care 24*, 2 (2012), 114–120.

[17] CHAPMAN, P., CLINTON, J., KERBER, R., KHABAZA, T., REINARTZ, T., SHEARER, C., AND WIRTH, R. CRISP-DM 1.0 Step-by-step data mining guide. Tech. rep., The CRISP-DM consortium, 2000.

[18] CLASSEN, D. C., PESTONIK, S. L., AND EVANS, R. S. Computerized surveillance of adverse drug events in hospital patients. *JAMA 266* (1991), 2847–2851.

[19] CUSTERS, T., ARAH, O. A., AND KLAZINGA, N. S. Is there a business case for quality in the netherlands?: A critical analysis of the recent reforms of the health care system. *Health Policy 82*, 2 (2007), 226–239.

[20] DBC-INFORMATIESYSTEEM. Ga sz 6.0 v 4.3: Standaard voor dis gegevensaanlevering dbc door aanbieders ziekenhuiszorg. Tech. rep., DIS, 2011.

[21] DBC-INFORMATIESYSTEEM. Website of the dutch dbc information system. `http://www.dbcinformatiesysteem.nl/`, 2011.

[22] DBC-ONDERHOUD. Website of the dutch dbc onderhoud. `http://www.dbconderhoud.nl/`, 2011.

[23] DBC ONDERHOUD. RSAD-document: Beschrijving van het RSAD-model, de basis van DOT, de nieuwe DBC-systematiek voor de ziekenhuiszorg. versie 20120329. Tech. rep., DBC Onderhoud, May 2012.

[24] DENZIN, N. K., AND LINCOLN, Y. S., Eds. *Strategies of Qualitative Inquiry*. SAGE, 2003, ch. 8: Grounded Theory (Kathy Charmaz), pp. 249–291.

[25] DM&IQ CONFERENCE EUROPE 2008. Iso 8000 - quality standards for data and information. Tech. rep., ISO/TC184/SC4/WG13/N0099 International Organization for Standardization, 2008.

[26] DOHMEN, A., AND LEYER, M. Combining data envelopment analysis and business process simulation for analyzing the efficiency of business processes. In *Proceedings of the International MultiConference of Engineers and Computer Scientists (IMECS) 2010* (Hong Kong, 2010), vol. 3.

[27] DUTCH HOSPITAL DATA. Website of dutch hospital data. `http://www.dutchhospitaldata.nl/`, 2011.

[28] EDITORS OF THE AMERICAN HERITAGE DICTIONARIES, Ed. *The American Heritage Medical Dictionary*. 2007.

[29] EDMONDSON, A. C., AND MCMANUS, S. E. Methodological fit in management field research. *Academy of Management Review 32*, 4 (2007), 1155–1179.

[30] EVANS, R. G. "behavioural"cost functions for hospitals. *Canadian Journal of Economics 4*, 2 (1971), 198–215.

[31] EVANS, S. M., BERRY, J. G., AND SMITH, B. J. Attitudes and barriers to incident reporting: a collaborative hospital study. *Qual Saf Health Care 15* (2006), 39–43.

[32] FARRELL, M. J. The measurement of productive efficiency. *Journal of the Royal Statistical Society 120*, 3 (1957), 253–281.

[33] FERNER, R. E., AND ARONSON, J. K. Clarification and terminology in medical errors: Definitions and classification. *Drug Safety 29*, 11 (2006), 1011–1022.

[34] FLICK, U. *An Introduction to Qualitative Research*. Sage, 2009.

[35] GALLAGHER, J. M., AND KUPAS, D. F. Experience with an anonymous web-based state ems safety incident reporting system. *Prehospital Emergency Care 16*, 1 (2011), 36–42.

[36] GENDRON, M. S., AND D'ONOFRIO, M. J. Data quality in the healthcare industry. *Data Quality 7*, 1 (2001).

[37] GHAURI, P., AND GRONHAUG, K. *Research methods in Business Studies*, 4 ed. Pearson Education Limited, Prentice Hall, Financial Times, 2010.

[38] GINNEKEN, E. V., SCHAFER, W., AND KRONEMAN, M. Managed competition in the netherlands: an example for others? *Eurohealth 16*, 4 (2010), 23–26.

[39] GOVERNMENT OF MANITOBA. The regional health authorities amendment and manitoba evidence amendment act - part 4.1: Patient safety. `http://web2.gov.mb.ca/laws/statutes/2005/c02405e.php`, 2005.

[40] HASAART, F. *Incentives in the Diagnosis Treatment Combination payment system for specialist medical care*. PhD thesis, Universiteit Maastricht, 2011.

[41] HELFERT, M., AND VON MAUR, E. A strategy for managing data quality in data warehouse systems. In *Proceedings of the Sixth International Conference on Information Quality* (Cambridge, 2001), E. M. Pierce and R. Kaatz-Haas, Eds.

[42] HELFERT, M., ZELLNER, G., AND SOUSA, C. Data quality problems and proactive data quality management in data warehouse systems. *Proceedings of the Business Information Technology World Conference* (2002).

[43] HELLINGS, J., VLEUGELS, A., SCHROOTEN, W., HAELTERMAN, H., PELEMAN, H., AND BORGERMANS, L. *Patintveiligheid in ziekenhuizen: basisinzichten en verkenning van veiligheidscultuur*. 2009.

[44] HOLLINGSWORTH, B. Non-parametric and parametric applications measuring efficiency in health care. *Health Care Management Science 6* (2003), 203–218.

[45] HOLLINGSWORTH, B., DAWSON, P. J., AND MANIADAKIS, N. Efficiency measurement of health care: a review of non-parametrics methods and applications. *Health Care Management Science 2* (1999), 161–172.

[46] HRIPCSAK, G., BAKKEN, S., STETSON, P. D., AND PATEL, V. L. Mining complex clinical data for patient safety research: a framework for event discovery. *Journal of Biomedical Informatics 36*, 1-2 (2003), 120–130.

[47] IEZZONI, L. I. Assessing quality using administrative data. *Ann intern med 127* (1997), 666–674.

[48] INSPECTIE VOOR DE GEZONDHEIDSZORG (IGZ). Invoering veiligheidsmanagementsysteem in ziekenhuizen komt op gang. Tech. rep., Den Haag, 2009.

[49] INSPECTIE VOOR DE GEZONDHEIDSZORG (IGZ). Veiligheidsindicatoren ziekenhuizen 2010-2012. toezicht op het veiligheidsprogramma "voorkom schade, werk veilig". Tech. rep., Den Haag, 2010.

[50] INSPECTIE VOOR DE GEZONDHEIDSZORG (IGZ). Staat van de gezondheidszorg 2011. Tech. rep., 2011.

[51] INSTITUTE OF MEDICINE. *To Err Is Human: Building a Safer Health System.* 1999.

[52] JHA, A. K., KUPERMAN, G. J., AND TEICH, J. M. Identifying adverse drug events: development of a computer-based monitor and comparison with chart review and stimulated voluntary report. *JAMA 5* (1998), 305314.

[53] KOOREMAN, P. Data envelopment analysis and parametric frontier estimation - complementary tools. *Journal of Health Economics 13*, 3 (1994), 345–346.

[54] KOOREMAN, P. Nursing home care in The Netherlands: a nonparametric efficiency analysis. *Journal of Health Economics 13* (1994), 301–316.

[55] KORNE, D. D. *Divergent Sight. Studies on the application of industrial quality and safety improvement methods in eye hospitals*. PhD thesis, University of Amsterdam, 2011.

[56] LANGELAAN, M., BAINES, R. J., BROEKENS, M. A., SIEMERINK, K. M., VAN DE STEEG, L., ASSCHEMAN, H., DE BRUIJNE, M. C., AND WAGNER, C. Monitor zorggerelateerde schade 2008. dossieronderzoek in nederlandse ziekenhuizen. Nivel, 2010.

[57] LEE, Y. W., STRONG, D. M., KAHN, B. K., AND WANG, R. Y. Aimq: a methodology for information quality assessment. *Information & Management 40* (2002), 133–146.

[58] LEGEMAATE, J., CHRISTIAANS-DINGELHOFF, I., DOPPEGIETER, R. M. S., AND ROODE, R. P. Melden van incidenten in de gezondheidszorg. Tech. rep., Onderzoek inm opdracht van ZonMw, 2006.

[59] LEITHEISER, R. L. Data quality in health care data warehouse environments. *Proceedings of the 34th Hawaii International Conference on System Sciences* (2001).

[60] LUDWIG, M. *Efficiency of Dutch hospitals*. PhD thesis, Universiteit Maastricht, 2008.

[61] MORTIMER, D., AND PEACOCK, S. Hospital efficiency measurement: Simple ratios vs. frontier methods. *Centre for health program evaluation: working paper 135* (2002).

[62] NICOLINI, D., WARING, J., AND MENGIS, J. Policy and practice in the use of root cause analysis to investigate clinical adverse events: mind the gap. *Social Science & Medicine 73*, 2 (2011), 217–225.

[63] ÖHRN, A., ELFSTRÖM, J., LIEDGREN, C., AND RUTBERG, H. Reporting of sentinel events in swedisch hospitals: a comparison of severe adverse events reported by patients and providers. *Joint Commission Journal on Quality and Patient Safety 37*, 11 (2011), 495–501.

[64] OKOLI, C., AND SCHABRAM, K. A guide to conducting a systematic literature reveiw of information systems research. *Sprouts: Working Papers on Information Systems 10*, 26 (2010).

[65] PIPINO, L. L., LEE, Y. W., AND WANG, R. Y. Data quality assessment. *Communications of the ACM 45*, 4 (2002), 211–218.

[66] REBASA, P., MORA, L., VALLVERDU, H., LUNA, A., MONTMANY, S., ROMAGUERA, A., AND NAVARRO, S. Adverse events in general surgery. a prospective analysis of 13,950 consecutive patients (article in spanish). *Cirugia Espanola 89* (2011), 599–605.

[67] RIGBY, K. D., AND LITT, J. C. B. Errors in health care management - what do they cost. *Quality in Health Care 9* (2000), 216–221.

[68] RUBEN, B. D. What patients remember: a content analysis of critical incidents in health care. *Health Communication 5*, 2 (1993), 99–112.

[69] RUNCIMAN, W. B., HELPS, S. C., SEXTON, E. J., AND MALPASS, A. A classification for incidents and accidents in the health-care system. *Journal of Quality in Clinical Practice 18*, 3 (1998), 199–211.

[70] SCHAAF, T. V. D. Prisma: a risk management tool based on incident analysis. In *Proceedings of the International Conference and Workshop on Process Safety Management and Inherently Safer Processes* (Orlando, Florida, 1996).

[71] SCHUT, F. T., AND VEN, W. P. M. M. V. D. Rationing and competition in the dutch health-care system. *Health Economics 14* (2005), 59–74.

[72] STELLA, J., BARTLEY, B., AND JENNINGS, P. Introduction of a prehospital critical incident monitoring system - final results. *Prehospital and Disaster Medicine 25*, 6 (2010), 515–520.

[73] STRAUSS, A. *Qualitative analysis for social scientists*. Cambridge University Press, 1987.

[74] TAYI, G. K., AND BALLOU, D. P. Examining data quality. *Communications of the ACM 41*, 2 (1998), 54–57.

[75] THE JOINT COMMISSION. Sentinel event policy and procedures. `http://www.jointcommission.org/Sentinel_Event_Policy_and_Procedures/`, 2011.

[76] THOMAS, E. J., AND PETERSON, L. A. Measuring errors and adverse events in health care. *J Gen Intern Med. 18* (2003), 61–67.

[77] VAN MONTFORT, G. P. W. P. Production functions for general hospitals. *Social Science Medicine. Part C: Medical Economics 15*, 2 (1981), 87–98.

[78] VINCENT, C., TAYLOR-ADAMS, S., CHAPMAN, E. J., HEWETT, D., PRIOR, S., STRANGE, P., AND TIZZARD, A. How to investigate and analyse clinical incidents: Clinical risk unit and association of litigation and risk management protocol. *British Medical Journal 320* (2000), 777–781.

[79] VINCENT, C., TAYLOR-ADAMS, S., AND STANHOPE, N. Framework for analysing risk and safety in clinical medicine. *Britisch Medical Journal 316* (1998), 1154–1157.

[80] VOGUS, T. J., AND SUTCLIFFE, K. M. The impact of safety organizing, trusted leadership, and care pathways on reported medication errors in hospital nursing units. *Medical Care 45*, 10 (2007), 997–1002.

[81] VUMC. Geconsolideerd maatschappelijk verantwoordingsdocument zorg - deel 1: Maatschappelijk verslag. Annual Report, 2010.

[82] VWS, IGZ, NVZ, NFU, ORDE, LEVV, AND V&VN. Voorkom schade, werk veilig. Tech. rep., 2007.

[83] WAGNER, C., CUPERUS-BOSMA, J., AND VAN DER WAL, G. Patient safety in dutch hospitals. *It J Public Health 2*, 3-4 (2005), 59–63.

[84] WAGNER, C., AND VAN DER WAL, G. Voor een goed begrip. bevordering patientveiligheid vraagt om heldere definities. *Medisch Contact 47* (2005), 1888–1891.

[85] WANG, R., AND STRONG, D. Beyond accuracy: what data quality means to data consumers. *Journal of Management Systems 12* (1996), 5–34.

[86] WANROOIJ, K. P. A. V. Patient Careflow Discovery. Master's thesis, Utrecht University, Utrecht, The Netherlands, 2012.

[87] WAUMANS, D. Patientveiligheid. Master's thesis, Universiteit Hasselt, 2008.

[88] WESTERDIJK, M., ZUURBIER, J., LUDWIG, M., AND PRINS, S. Defining care products to finance health care in the netherlands. *The European Journal of Health Economics* (2011), 1–19.

[89] WIRTH, R., AND HIPP, J. Crisp-dm: Towards a standard process model for data mining. In *Proceedings of the Fourth international Conference on the Practical Applications of Knowledge Discovery and Data Mining* (Manchester, UK, 2000), pp. 22–39.

[90] WONG, C. A., RECKTENWALD, A. J., JONES, M. L., WATERMAN, B. M., BOLLINI, M. L., AND DUNAGAN, W. C. The cost of serious fall-related injuries at three midwestern hospitals. *Joint Commission Journal on Quality and Patient Safety 37* (2011), 81–87.

[91] WORTHINGTON, A. C. Frontier efficiency measurement in healthcare - a review of empirical techniques and selected applications. *Medical Care Research and Review 61*, 2 (2004), 1–36.

[92] YIN, R. K. *Case Study Research: Design and Methods*. 2003.

[93] ZEGERS, M., BRUINE, M. C., WAGNER, C., GROENEWEGEN, P. P., WAAIJMAN, R., AND WAL, G. V. D. Design of a retrospective patient record study on the occurrence of adverse events among patients in dutch hospitals. *BMC Health Services Research 7*, 27 (2007).

Part IV

Appendices

A Interview Protocol

Written by: Peter van Stijn
Date: April 2012

This protocol was created before the interviews took place. During the interviews, a combination of both this protocol, as the document "background information for interviewees" (Appendix B) were used as guideline. The background information was created in a later phase and is in some places more up to date.

A.1 Introduction

This document provides the protocol that is used to structure the expert interviews performed in this Master's Thesis. The protocol is based on the descriptions provided by Yin [92]. Although those descriptions are primarily aimed at designing a *Case Study Protocol*, interviews play an important role. The aim of the protocol is to structure the interviews in such a way to enable comparison of the different interview results. Most scientific references to concepts and methodologies used in this protocol are provided in the research method and literature sections (Sections 2 and 4) in this Thesis. This protocol serves as a more detailed extension of these sections.

This document is structured as follows: Section A.2 will start with repeating the thesis research questions and providing a description of the topics for which the expert interviews will serve as a main data acquiring method. Section A.3.1 describes the interviewees, their relation with the topics, and their relation with each other. Section A.3.3 provides an overview of the secondary sources that are used within the interviews. The interview outline is structured in Section A.4. *This outline should was, during the interviews, combined with the background information for interviewees (Appendix B) document and the constructed data quality framework*.

A.2 Interview topics

The Main Research Question of this Master Thesis has been formulated and split in separate sub questions as follows (see also Section 1.2 of this thesis):

- *"What is the Quality of the DIS data and its usefulness in the context of efficiency and patient safety measurements, and what organizational and technical factors influence this quality?"*
    1. How can data quality be measured?
    2. How can (hospital) efficiency be specified?
    3. How can patient safety be specified?
    4. What elements do the DIS data contain and how do these relate?
    5. How are the DIS data obtained, from the moment of registration to the final database insertion?
    6. What is the Intrinsic Quality of the DIS data and what organizational and technical issues influence this?
    7. What is the Contextual Quality of the DIS data with regard to evaluating efficiency?

8. What is the Contextual Quality of the DIS data with regard to evaluating patient safety?
9. How can the DIS data quality be improved?

As described in section 2 of this Thesis, the questions 4 to 8 are answered with the results from the expert interviews. These five sub questions result in three main interview topics: the Data Quality, Efficiency, and Patient Safety. For each interviewee, the focus of the interview is on one of these specific topics, although their views on the other topics are assessed as well. The data itself (sub section 4) are being discussed in all these topics, and within the topic of data quality, sub questions 5 and 6 are assessed.

A.2.1 Data quality

The first goal of the interviews is to determine the quality of the DBC data, as this influences the reliability of the (efficiency and safety) conclusions that are drawn from the data [7]. A lot has been written on data quality measurement, although mostly in combination with methodologies on improving this data quality. This research focuses only on measuring the quality. Many different data quality dimensions have been identified through literature, with many possible categorizations. A very useful categorization is the one from the data quality framework provided by Wang and Strong [85], which has also been evaluated in a hospital setting [36]. This framework, constructed in an empirical approach from 179 possible dimensions, categorizes 15 (originally 20) dimensions in four categories: accuracy, relevancy, representation and accessibility. Batini and Scannapiena rename the accuracy dimension to *Intrinsic*, and the relevancy dimension to *Contextual* [7]. In their book, they also describe the different focus of the four dimensions:

The *Intrinsic* data quality describes the quality of the data itself, while the *Contextual* data quality considers the quality related to the context in which the data is used. Furthermore, the category *Representational* data quality exists, which considers aspects such as conciseness, interpretability and ease of understanding. Finally, *Accessibility* data quality considers properties of data access, such as accessibility itself and access security [7, 85].

For this Master's Thesis, the latter two categories (i.e., representational and accessibility) are out of scope and will not be investigated further. The second category is in this thesis scoped to the contexts of efficiency and patient safety measurement. The first category (intrinsic quality) is important to determine the reliability of the conclusions drawn from the data. The data quality framework developed by Wang and Strong provides the following dimensions for these two categories [85], where the average importance in a health care setting [36] is noted between the brackets (scale 1 to 5, 1 represents extremely important): (note, this leaves open the possibility to not investigate "value-added" and "appropriate amount of data")...

- Intrinsic Data Quality
  1. *Believability* (1.73): believable
  2. *Accuracy* (1.47): data are certified error-free, accurate, correct, flawless, reliable, errors can be easily identified, the integrity of the data, precise
  3. *Objectivity* (1.87): unbiased, objective
  4. *Reputation* (1.97): the reputation of the data source, the reputation of the data

- Contextual Data Quality
    1. *Value-added* (2.06): data provide competitive advantage, data add value to operations
    2. *Relevancy* (1.89): applicable, relevant, interesting, usable
    3. *Timeliness* (1.91): age of data
    4. *Completeness* (1.84): the breath, depth and scope of information contained in the data
    5. *Appropriate amount of data* (2.09): the amount of data is appropriate to the task at hand

For the dimension *reputation*, especially related to the reliability of the source, it is also important to sketch and analyze the path the data goes through from DBC registrations towards DIS delivery. These data edits and transactions are prescribed, but it is not unlikely hospitals differ from this prescribed path.

Batini and Scanapieca [7] have provided a further separation of the dimensions accuracy, timeliness and completeness. For accuracy, semantic and syntactic accuracy are recognized. For timeliness; currency, volatility and timeliness (age) itself are recognized. Finally, for completeness; column, schema and population completeness are differentiated. During the interviews, this further separation into sub-dimensions will be used to assess the dimensions in more detail and have clear understanding on the scope of these dimensions. Furthermore, except for the sub-dimensions schema completeness and timeliness (age), the ”new” sub-dimensions are moved to the intrinsic data quality dimension.

A.2.2 Efficiency

The first context to measure the contextual data quality in, is (department) efficiency. The background literature subsection 4.2 describes the concepts related to efficiency and measurement methods in more detail.

Mainly, three different types of efficiency exist: *technical efficiency*, which relates to the rate of output that is provided from a certain amount of input; *allocative efficiency*, which relates to the rate in which the proportions of the used inputs (to produce an output) are in optimal proportions; and *economic efficiency*, which combines technical and allocative efficiency: the rate in which a company produces a maximal output from a cost-minimizing input mix.

Different approaches exist to measure (different types of) efficiency. These differ from parametric versus non-parametric approaches; deterministic versus stochastic approaches; average versus frontier approaches; and approaches using multi-variate versus bi-variate indicators (e.g., input/output ratio versus more complex cost functions). The approaches used mostly in a hospital setting are SFA (a parametric, stochastic approach) and DEA (a non-parametric, deterministic approach): both of these approaches are multi-variate and frontier approaches. Other methods used in health care settings are DFA (SFA but deterministic); MI (DEA measuring efficiency over time); Regression Analysis (like SFA parametric and stochastic, but measuring efficiency against an average instead frontier); and simple ratio efficiency (simple bi-variate cost/output ratios).

As described often in literature, the main problem with efficiency measurements in a hospital setting is quantifying the outputs, as improved health status or quality of life is difficult to measure. Often, measurements of physical performance such as patient days are used. Specifying inputs (or, constructing cost functions) is also not straightforward considering the many possible input factors (such as used personnel), although less complex than specifying outputs as these input factors are by definition more quantitative.

Whether the DBC data provides enough information to measure these inputs and outputs, is the main question of this interview topic.

A.2.3 Patient Safety

The second context to measure the contextual data quality in, is (department) patient safety. The background literature subsection 4.3 describes the concepts related to safety and (incident) measurement methods (especially registration and analysis) in more detail.

Patient safety refers to the risk a patient has of suffering any damage caused by health care professionals, which could have been prevented. This means the occurrence of incidents, and especially adverse events, has to be minimized. The definition brings up the necessity of defining a number of related concepts:

An *incident* is an unintended happening that has led, could have led, or could still lead to injury to the patient [87]. These incidents have occurred during or following a medical intervention and are thus iatrogenic: "Induced in a patient by a physician's activity, manner, or therapy" [28, 43]; or induced by health care management in general (e.g., the information systems) [93]. Some incidents are the result of a (human) failure in the treatment path: those incidents are referred to as *Medical Errors*. Neither incidents, nor medical errors *all* result in injury. Those that do, are called *Adverse Events* [87, 93]. If an adverse event resulted from a medical error (thus, a subgroup of adverse events), it is called an *Avoidable Adverse Event* [51]. Incidents that did not result in injury are referred to as *Near Misses*. All adverse events are also referred to as *Complications*, but these complications are a wider concept: e.g., also unintended happenings resulting from the patients primary disease or calculated risks are included [83]. The concepts are described in more detail in Section 4.3.1, and the relation between an incident, medical error, adverse event and avoidable adverse event is visualized in Figure 13.

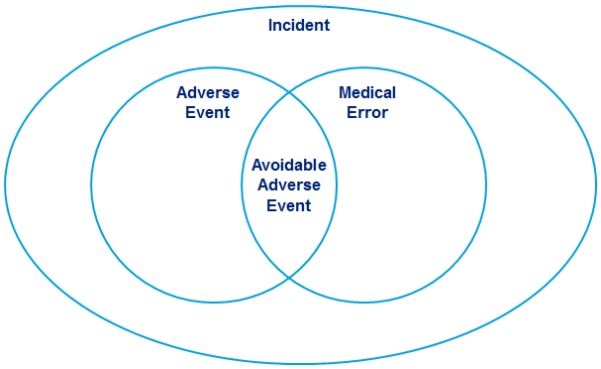

Figure 10: Venn Diagram showing the relationship between the different concepts

The main question of this interview topic is whether the DBC data, which includes information on the operations performed on a patient, provides enough information to identify incidents, and thus patient safety.

A.3 Data Collection Procedures

This section describes what data will be collected from what interviewees, and how these data will be used in the Thesis. It starts with a description of the interviewees in section A.3.1. The section also provides information on how these experts relate to each other and to which extent their input will be of value in the research. Section A.3.3 describes the external sources that will be used directly in the interviews.

A.3.1 Interviewees

Regarding the hospital (DBC) data, multiple stakeholders are involved. These are (employees from) the hospitals themselves, or more specifically the hospital health administration departments; the health insurers; the government; DBC Onderhoud (including the DIS: Dutch DBC Information System); Consultants directly employed by the hospital; External consultants hired by the hospital; and scientists. The roles of these different stakeholders are described below:

Hospital Health Administration Department These departments are usually called *zorgadministratie* (health care administrations) and are responsible that the data (DBC registrations) is entered correctly into the systems, such that these data can sufficiently be used and analyzed on the back-end. For this research, employees from this administration department will be very useful, especially to cover the topic of data quality. In order to ensure experience and a good overview of these aspects, this employee will have to be of the senior level (e.g., the head of department or senior manager).

Hospital Business Intelligence Hospital Business Intelligence employees are used to perform analyses on hospital data and often also perform certain analyses on the DBC data, although only when still within the hospital. Furthermore, there is often a financial focus of the analyses performed. These stakeholders will be very useful to provide more insights on the contextual data quality of DIS, especially in relation to efficiency measurements.

Health Insurers The hospitals gather the data and provide them to the health insurers with the purpose of reimbursement. The data on the provided care also serves as input for future reimbursement (prices) negotiations. For this research, employees from insurance companies will not be used (for that specific stakeholder's view). The reason for this is that insurers do not get the detailed DBC data but merely summaries (for reimbursement purposes) and thus can not evaluate the quality of these data. Furthermore, as these insurers do not have specific knowledge on patient safety or efficiency (from that role), they seem less useful to serve as input for those topics as well.

Government Although not directly involved with the data, the Dutch government "manages" the *managed competition* which exists in the Dutch health sector. For experts representing this segmentation, the same counts as for insurers: these experts will not be used in the interviews for this Thesis.

DBC Onderhoud DBC onderhoud (DBC maintanance) is the Dutch expertise- and servicecentre related to the DBC system responsible for the collection, validation, governance and distribution of the DBC data. It consists of four main departments: *Releases*, which specify, develop, test and

deliver DBCs; *Servicecentre*, which serves as a helpdesk; *Staff services* of DBC Onderhoud itself; and *DBC-Information System (DIS)*. The latter governs all the data from the DBCs. [22]. For this thesis, experts representing DBC Onderhoud will be very useful, especially on the topic of data quality. The departments of releases, but especially DIS itself, will be most useful to

Internal Consultants Consultants directly employed by the hospital to perform various analyses or preparations on the data;

External Consultants External consultants hired by the hospital to analyse or modify these data... Internal or External (health care) consultants are often consulted to analyze the gathered hospital data (e.g., for DOT impact analysis) or to help preparing these data for the provision to DIS.

Scientists Scientists do not focus on the gathered data itself, but usually have extensive knowledge on a specific field (e.g., hospital efficiency or patient safety). This knowledge can be assessed to analyze the usability of the data.

For this research, the potential interviewees are determined using purposive sampling, where participants representing specific characteristics are purposively selected [13].

A.3.2 Sampling, Stakeholders

Refer to Appendices for more information: Interview Protocol. Describes the selection of the different stakeholders and the interviewees.

Flick [34, p. 123] describes that the different aims of sampling (width, depth) of the research, considering the limited resources (people, money, time, etc.) should been seen as alternatives rather than projects to combine. For this research and the corresponding expert interviews, three topics can be identified: determining the DIS data quality; finding determinants from these data to describe efficiency; and finding determinants from these data to describe patient safety.

For the data quality assessment, the weighting between width (to study the data quality from as many different health institutions as possible) and depth (to study the data quality extensively for university hospitals) resulted in a more in-depth research. Only the data quality of university hospitals (only health care, not research and education) is examined. However, to ensure some width (or, internal width of the chosen field would be a better term), multiple stakeholders in this data quality are interviewed. First, the information managers within the hospitals who are responsible for registration of the data; second, the managers business intelligence within the hospital who are responsible for the internal analyses of the data; and third, the managers or employees from DBC Onderhoud and DIS, who receive the data and validate, store, analyze and distribute the data further.

For the contextual quality of the data to describe efficiency or patient safety, the width dimension was more important. As the research on these sub-questions is more explorative, it is important to cover insights from as many stakeholders as possible on these aspects.

The experts that will be interviewed will each represent different stakeholders (with corresponding insights) towards the topic. These different ”invalshoeken” are the following:

Internal or External (health care) consultants are often consulted to analyze the gathered hospital data (e.g., for DOT impact analysis) or to help preparing these data for the provision to DIS.

For data quality assessment: two of the stakeholder groups are most important:
- Hospital Information Managers (e.g., hoofd zorgadministratie): the experts responsible for the gathering, editing and delivery of the data
- DBC onderhoud / DIS: the experts that receive the data and combine it to the central database

Furthermore, consultants (both working externally and internally to hospitals) are interviewed on the fields of either efficiency and safety: these are not as closely related to the data but there is no doubt they can say something on it.

For efficiency: the internal consultants, external consultants, and scientists are most important. However, the experts working more closely to the data (knowing what is in it better) might also prove very useful to say something on this efficiency.

For safety: the internal consultants, external consultants, and scientists are most important. However, the experts working more closely to the data (knowing what is in it better) might also prove very useful to say something on this safety.

Different dimensions represent different stakeholders: those responsible for the data generation, those receiving the data, those working (internally and externally) with the data and scientists. An expert represents the opinion of a group of stakeholders; at least one expert for each focus area (data quality, efficiency, safety)

A.3.3 Secondary Sources

This section shortly describes the sources to take to the interview, which can serve at certain points in time as reference work:

DIS aanleverformaat ziekenhuiszorg

ERD diagram DBC data page 12 and 13 form the DIS aanleverformaat, printed on separate paper to make notes.

beschrijving aanleverproces

Thesis version 0.52

Kladblok for notes

Laptop Turned in sleep modus, for references if needed

Phone with Voice recording All interviews will be recorded.

A.4 Interview Outline (Topic List; Dutch)

A.4.1 Introductie

Introduction of the interview. Max 15 minutes.

1. Introductie to self. Study, previous study, etcetera.
2. Explanation of the research.
3. Ask the interviewee to introduce him/herself.

A.4.2 DBC data kwaliteit

1. Assessment bekendheid interviewee met topic data kwaliteit, en bekendheid interviewee met data zelf.
2. data path assessment: explanation of prescribed path, what is the role of the interviewee in the path, explanation of the real path
3. Quality indicators: Grade relevancy of indicators 1-10 (10 best). Then, grade dimensions themselves 1-10 (10 best) and explain further.
    - Explanation intrinsic data quality, context, representational, accessibility; scope
    - Accuracy:
        - Syntactic Accuracy
        - Semantic Accuracy
    - Time-Related Dimensions:
        - Currency
        - Volatility
        - (Timeliness)
    - Completeness:
        - Column Completeness
        - Population Completeness
    - Believability
    - Objectivity (Unbiased)
    - Reputation (data and data sources). Who enters the data? When?

A.4.3 Het afleiden van efficiëntie uit de DBC data

Eerst kort kijken naar de verschillende efficiënties die vaak worden gemeten en hoe we er bij voorbaat over denken. Denk de verschillende input en output factoren die normaal worden gebruikt bij ziekenhuizen vergelijken met wat mogelijk is met deze data. Welke gegevens mis je en hoe ernstig is dit? Welke gegevens heb je extra tot je beschikking en hoe bruikbaar is dit? Tot slot kijken naar verschillende efficiëntie meetmethoden en of die geschikt zijn in combinatie met deze data.

1. Assessment bekendheid interviewee met topic
2. technische, allocatieve, economische efficiëntie.
    - Is het mogelijk technische efficiëntie af te leiden uit de data? (Maximum output van vaste input, of minimale input voor vaste output).
    - Is het mogelijk voor allocatieve efficiëntie? (optimale input mix).
    - Eventueel de combinatie, economische efficiëntie?
3. Bepalen inputfactoren en output factoren vanuit data (versus normaal gesproken)
    - Veelgebruikte inputfactoren in onderzoeken: capital (volgens Ludwig PhD niet nodig mee te nemen in Nederland); Staff (prices, 60% van kosten); Materials (40% totale kosten). Hoe ziet u dit? Missen hier factoren?
    - Veelgebruikte outputfactoren in onderzoeken: Patient days; (gewogen) (dag-)opnames; Ludwig PhD laat research weg, maar bekijkt wel voor onderwijs ”number of trained specialists”.
    - Output vanuit DBC perspectief: puur financieel bekeken, het geld dat is gegenereerd met de (DBC) handelingen, dus de reimbursements van de verzekeraar. Input zijn dan die handelingen, of specifieker de totale kosten van die handelingen. Wat is de mening van de interviewee wat met de DBC data voor input/output waarden kan bekeken, of eventueel om te gebruiken voor het opstellen van de kostfunctie(s)?
    - Is de DBC systematiek wat dit betreft specialisme overstijgend?
4. Meet methodieken
    - bekendheid interviewee met methodieken als DEA en SFA
    - Parametrisch versus non-parametrisch ; Stochastisch versus deterministisch ; Average versus frontier; multi-variate versus bi-variate (?)
5. Quality indicators: Grade relevancy of indicators 1-10 (10 best). Then, grade dimensions themselves 1-10 (10 best) and explain further.
    - Value Added: Can these data provide competitive advantage or add value to operations?
    - Relevancy: Is de data toepasbaar, relevant, bruikbaar?
    - Timeliness: Is de ”leeftijd” (age) van de data goed genoeg om conclusies van te trekken? Is dit relevant?
    - Completeness: De breedte, diepte en scope van de informatie in de data: is die compleet genoeg om efficiëntie af te leiden?
    - Appropiriate Amount: Is de hoeveelheid data geschikt? Is dit relevant?

A.4.4 Het afleiden van patiëntveiligheid uit de DBC data

1. Assessment bekendheid interviewee met topic
2. Explanation Incident, Adverse Event, Medical Error, Preventable Adverse Event, Complication.
3. Huidige maatstaven van patiëntveiligheid (enkel afleiden uit vóórkomen incidenten)
4. Data in VMS (al is het niet gestandaardiseerd). Wat is Wat zit eigenlijk ook al in DBC? Wat mist er?
5. Performance Indicators patiëntveiligheid mogelijk vanuit DIS
6. Welke aspecten missen bij de DBC registratie data.
7. Quality indicators: Grade relevancy of indicators 1-10 (10 best). Then, grade dimensions themselves 1-10 (10 best) and explain further.
    - Value Added: Can these data provide competitive advantage or add value to operations?
    - Relevancy: Is de data toepasbaar, relevant, bruikbaar?
    - Timeliness: Is de ”leeftijd” (age) van de data goed genoeg om conclusies van te trekken? Is dit relevant?
    - Completeness: De breedte, diepte en scope van de informatie in de data: is die compleet genoeg om veiligheid (incidenten) af te leiden?
    - Appropiriate Amount: Is de hoeveelheid data geschikt? Is dit relevant?

A.4.5 Afsluitende vragen

1. Zijn er nog dingen die u tijdens dit interview niet kwijt hebt gekunt, maar die in deze context wel heel belangrijk zijn te vermelden?

B Background Information for the Experts

B.1 Inleiding

Bedankt voor uw medewerking aan mijn Master thesis. Van een aantal personen kreeg ik het verzoek wat achtergrond informatie en/of onderzoeksvragen van de thesis te krijgen. Om die reden heb ik dit document opgesteld.

Het document begint met een korte beschrijving van de thesis, de achterliggende gedachte en de resultaten die ik voor ogen heb. Vervolgens worden de hoofdvraag en subvragen (Sectie B.3) gegeven, zoals ze op dit moment geformuleerd zijn. Dan volgt voor ieder thesis onderdeel (datakwaliteit, efficiëntie en patiëntveiligheid) een korte beschrijving met de belangrijkste aspecten. Over het algemeen wordt bij de interviews slechts focus op één van deze onderdelen gelegd, afhankelijk van de geïnterviewde. De onderdelen die tijdens het interview niet of nauwelijks behandeld zullen worden, kunnen wat dat betreft gerust overgeslagen worden. De onderzoeksvragen en samenvattingen van de drie onderzoekstopics zijn in het Engels geschreven, omdat de voertaal van de scriptie ook Engels is en hiermee ook verwarring in concepten zoveel mogelijk wordt voorkomen. Vrijwel alle statements en concept definities zijn gebaseerd op literatuur, maar dit is voor deze samenvatting niet opgenomen. U zult van het uiteindelijke thesis ook een kopie ontvangen, daar is uiteraard wel overal netjes gerefereerd.

B.2 Korte beschrijving thesis

Mijn scriptie richt zich op de data zoals die gestandaardiseerd door alle ziekenhuizen aan het DIS (DBC Informatiesysteem) aangeleverd wordt. Deze data bestaat voornamelijk uit DBC data: de verrichtingen uitgevoerd op een bepaalde patiënt, binnen een bepaalde DBC, binnen een bepaald zorgpad. Daar deze data geprotocolleerd is (en daar ook op wordt gevalideerd), kunnen we er vanuit gaan dat we te maken hebben met uniforme data. Hierdoor leent deze data zich goed voor benchmarks tussen ziekenhuizen onderling.

De benchmarks die ik hierbij voor ogen heb richten zich op efficiëntie en patiëntveiligheid (als onderdeel van zorgkwaliteit). Een van de achterliggende gedachtes achter de DBC systematiek was natuurlijk dat de efficiëntie van de verleende zorg zou stijgen (vanwege vaste vergoedingen voor gehele DBCs ongeacht de verrichtingen). Daarnaast zou de zorgkwaliteit (met patiëntveiligheid als belangrijk aspect hiervan) gewaarborgd moeten blijven of zelfs verbeteren, omdat de zorgverzekeraars (uiteindelijk) selectief zorgproducten in zouden kopen bij de aanbieders. Daarnaast geeft dit financiële systeem veel aanknopingspunten voor efficiëntieberekeningen, en is het een interessante vraag of patiëntveiligheid (of het vóórkomen van incidenten) afgeleid kan worden uit de manier waarop een patiënt is geholpen.

Alvorens analyses op de DIS data los te laten is het belangrijk op de hoogte te zijn van de beperkingen van deze data, of algemeen gezegd de data kwaliteit. Het eerste deel van mijn onderzoek richt zich dan ook hierop. Hierin is het aanleverproces van basisregistraties tot uiteindelijk het DIS ook heel interessant. De verschillende stapjes, bijvoorbeeld het halen van de DBC data door de grouper, zullen invloed hebben op deze datakwaliteit. Het goed in kaart brengen van dit aanleverproces, dat geschetst is in Figuur 11, en het beoordelen van de verschillende kwaliteitsdimensies is dan ook het eerste deel van mijn onderzoek.

B.3 Onderzoeksvragen

De onderzoeksvragen zoals ze op dit moment in het Engels opgesteld zijn voor de thesis, luiden als volgt:

- *"What is the quality of the DIS data from Dutch academic hospitals to estimate efficiency and patient safety, and what is the relation between these aspects on the level of hospitals and their divisions?"*

    1. How can efficiency within hospital departments be specified?
    2. How can patient safety be specified?
    3. What has been researched on the relation between efficiency and patient safety?
    4. What elements do the DIS data contain and what is the quality of these data?
    5. How useful is the DIS data from Dutch academic hospitals to evaluate efficiency?
    6. How useful is the DIS data from Dutch academic hospitals to evaluate patient safety?
    7. What can be said on the relation between efficiency and patient safety, based on the DIS data from Dutch academic hospitals?

De vragen 4, 5 en 6 worden met behulp van expert interviews beantwoord. Mocht het uiteindelijk niet mogelijk blijken ook daadwerkelijk analyses toe te passen op de data, zal vraag 7 worden "weggescoped".

B.4 DIS data

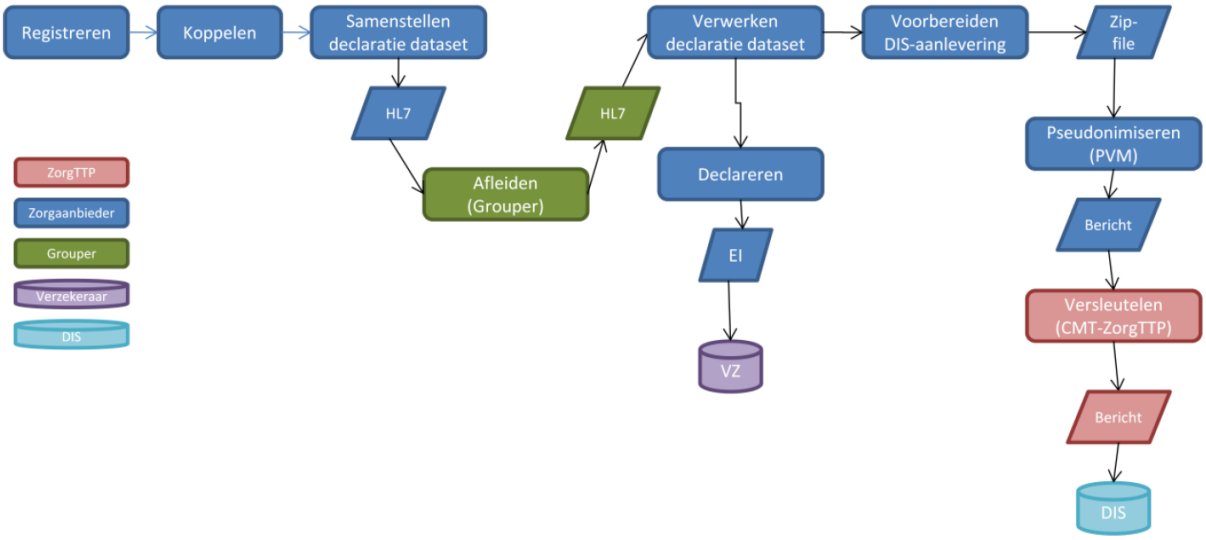

Figure 11: Aanleverproces DBC data naar DIS. (bron: standaard voor DIS gegevensaanlevering)

Nadat de ziekenhuizen hun declaraties hebben gedaan bij de verzekeraars, sturen ze deze (DBC) data door naar het landelijke DIS, het DBC Informatiesysteem. Hiervoor is een uitgebreide standaard voor de gegevensaanlevering opgesteld.

Figuren 11 en 12 zijn onderdeel van deze standaard en geven respectievelijk een schematisch beeld van het aanleverproces weer, en een samenvatting van de tabelstructuur met de kolommen die hierin voorkomen.

De data die ik uiteindelijk zou willen analyseren betreft de data na validatie door het DIS. Dus de data wat de ziekenhuizen volgens de standaard hebben aangeleverd, wat vervolgens is gecontroleerd (gevalideerd en geflagd) door het DIS en daarna in hun databases is opgeslagen. Deze data wordt op dit moment (al dan niet beknopt) opgeleverd aan wettelijk rechthebbende partijen zoals de NZA en CBS.

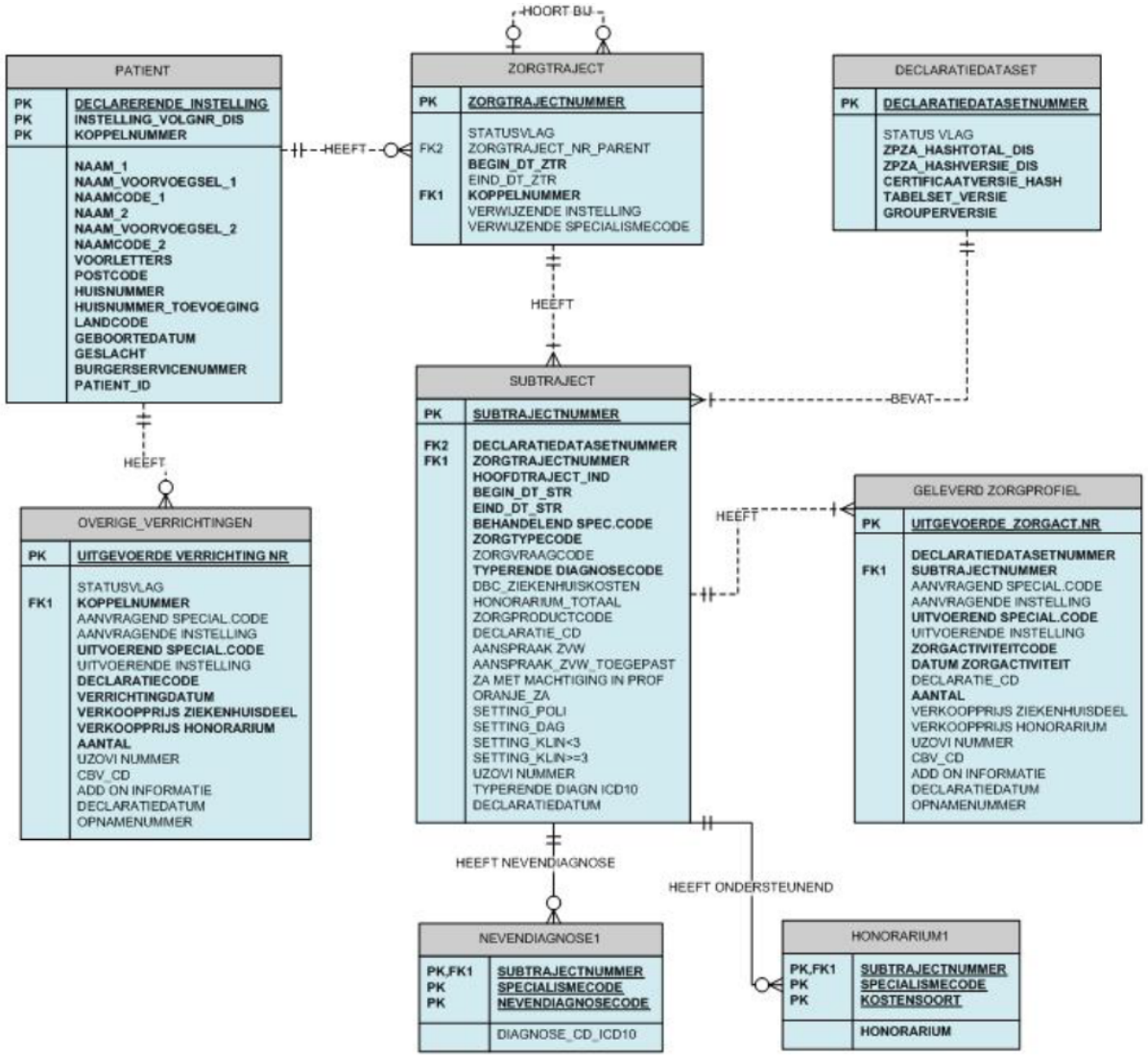

Figure 12: Samenvatting van DIS datastructuur (bron: standaard voor DIS gegevensaanlevering)

B.5 Onderzoekstopics (English summary)

B.5.1 Data quality

The first goal of the interviews is to determine the quality of the DBC data, as this influences the reliability of the (efficiency and safety) conclusions that are drawn from the data. A lot has been written on data quality measurement, although mostly in combination with methodologies on improving this data quality. This research focuses only on measuring the quality. Many different data quality dimensions have been identified through literature, with many possible categorizations. A very useful categorization is the one from the data quality framework provided by Wang and Strong, which has also been evaluated in a hospital setting. This framework, constructed in an empirical approach from 179 possible dimensions, categorizes 15 (originally 20) dimensions in four categories: accuracy, relevancy, representation and accessibility. Batini and Scannapiena rename the accuracy dimension to *Intrinsic*, and the relevancy dimension to *Contextual*. In their book, they also describe the different focus of the four dimensions:

The *Intrinsic* data quality describes the quality of the data itself, while the *Contextual* data quality considers the quality related to the context in which the data is used. Furthermore, the category *Representational* data quality exists, which considers aspects such as conciseness, interpretability and ease of understanding. Finally, *Accessibility* data quality considers properties of data access, such as accessibility itself and access security.

For this Master's Thesis, the latter two categories (i.e., representational and accessibility) are out of scope and will not be investigated further. The second category is in this thesis scoped to the contexts of efficiency and patient safety measurement. The first category (intrinsic quality) is important to determine the reliability of the conclusions drawn from the data. The data quality framework developed by Wang and Strong provides the following dimensions for these two categories (translated to Dutch for interview purposes):

- Intrinsieke Data Kwaliteit
    1. *Accuraatheid*:
        (a) *Syntactische Accuraatheid*: Klopt de data qua vorm? Bijvoorbeeld, staat er in het veld "activity" een bestaande activitycode? Is een naam een echte naam?
        (b) *Semantische Accuraatheid*: Klopt de data qua inhoud? Bijvoorbeeld, is de gegeven activitycode ook echt de activity die is verricht? Is de naam de juiste naam?
    2. *Tijdigheid*:
        (a) *Currency*: Hoe snel wordt de data in het systeem gevoerd (bij registratie), nadat de acties (verrichtingen) plaats hebben gevonden?
        (b) *Timeliness*: Hoe oud is de data? (en dan hoe oud dit is waarvoor je het wilt gebruiken en hoe ernstig leeftijd daarin is).
        (c) *Volatility*: Hoe vaak wordt de data geüpdatet? Bijvoorbeeld, bij een geboortedatum verwacht je geen updates, maar wijzigingen in eerder aan het DIS opgeleverde DBC data?
    3. *Compleetheid*:
        (a) *Column Completeness*: Bijvoorbeeld missing values.

(b) *Population Completeness*: Staan alle patiënten, DBCs en/of verrichtingen in de database?

(c) *Schema Completeness*: De laatste compleetheid dimentie heeft ook te maken met de context (quality). Schema compleetheid gaat over de volledigheid van het relationele schema, dus het datadiagram (Figuur 12), voor de taak waar je het voor wilt gebruiken. Mis je essentiële data voor het uitvoeren van bijvoorbeeld efficiëntieberekeningen?

4. *Geloofwaardigheid*: Is de data geloofwaardig, in die zin dat het geloofwaardig is dat hetgeen dat erin zit representatief is voor wat er werkelijk is gebeurd? Dus eigenlijk het antwoord op de vraag in hoeverre mensen er baat bij hebben andere data in te voeren dan wat er werkelijk is gebeurd (bv upcoding).
5. *Objectiviteit*: Is de data objectief (gebaseerd op feiten) of subjectief (gebaseerd op interpretaties)?
6. *Reputatie*: Dit slaat voornamelijk op de reputatie van de databron: wie voert de data in en is die bekend met de (semantische) inhoud van deze data?

For the dimension *reputation*, especially related to the reliability of the source, it is also important to sketch and analyze the path the data goes through from DBC registrations towards DIS delivery. These data edits and transactions are prescribed, but it is not unlikely hospitals differ from this prescribed path.

To measure the data quality, first the path of the data (from registration to DIS) will be sketched, including the steps that definately will influence data quality. Then, this data quality will be split into the multiple dimensions, and each dimension will be discussed separately with the interviewee. This includes a short discussion on the aspect itself and some examples or anekdotes to clarify this in the specific hospital. Furthermore, the interviewee will be asked to grade the dimension on a scale 1 to 10 (1=worst).

B.5.2 Efficiency

The first context to measure the contextual data quality in, is (department) efficiency. The background literature subsection describes the concepts related to efficiency and measurement methods in more detail.

Mainly, three different types of efficiency exist: *technical efficiency*, which relates to the rate of output that is provided from a certain amount of input; *allocative efficiency*, which relates to the rate in which the proportions of the used inputs (to produce an output) are in optimal proportions; and *economic efficiency*, which combines technical and allocative efficiency: the rate in which a company produces a maximal output from a cost-minimizing input mix.

Different approaches exist to measure (different types of) efficiency. These differ from parametric versus non-parametric approaches; deterministic versus stochastic approaches; average versus frontier approaches; and approaches using multi-variate versus bi-variate indicators (e.g., input/output ratio versus more complex cost functions). The approaches used mostly in a hospital setting are SFA (a parametric, stochastic approach) and DEA (a non-parametric, deterministic approach): both of these approaches are multi-variate and frontier approaches. Other methods used in health care settings are DFA (SFA but deterministic); MI (DEA measuring efficiency over time); Regression Analysis (like SFA parametric and stochastic, but measuring efficiency against an average instead frontier); and simple ratio efficiency (simple bi-variate cost/output ratios).

As described often in literature, the main problem with efficiency measurements in a hospital setting is quantifying the outputs, as improved health status or quality of life is difficult to measure. Often, measurements of physical performance such as patient days are used. Specifying inputs (or, constructing cost functions) is also not straightforward considering the many possible input factors (such as used personnel), although less complex than specifying outputs as these input factors are by definition more quantitative.

Whether the DBC data provides enough information to measure these inputs and outputs, is the main question of this interview topic. To answer the question, the data model will be looked at. What different efficiency related aspects are stored in the data, and are this useful for calculations? What calculations would be possible, and what would the answer state? Can, for example, cost prices be extracted from this DIS data directly? Where do the prices come from and how are they calculated? Would a calulation of efficiency based on total costs of operations (with one DBC), measured against the total reimbursement from insurers for that specific DBC be accurate? Is it a good idea to take the reimbursements as output? What information do we need further (miss) for such calculations? Are efficiency measurements used more often in hospital settings, such as SFA and DEA, possible and suited in this context? Can you say something on the following quality indicators that refer to the contextual quality of the data?

- Value Added: Can these data provide competitive advantage or add value to operations?
- Relevancy: Is de data toepasbaar, relevant, bruikbaar?
- Timeliness: Is de ”leeftijd” (age) van de data goed genoeg om conclusies van te trekken?
- Completeness: De breedte, diepte en scope van de informatie in de data: is die compleet genoeg?
- Appropiriate Amount: Is de hoeveelheid data geschikt? Is dit relevant?

B.5.3 Patient Safety

The second context to measure the contextual data quality in, is (department) patient safety. The background literature subsection describes the concepts related to safety and (incident) measurement methods (especially registration and analysis) in more detail.

Patient safety refers to the risk a patient has of suffering any damage caused by health care professionals, which could have been prevented. This means the occurrence of incidents, and especially adverse events, has to be minimized. The definition brings up the necessity of defining a number of related concepts:

An *incident* is an unintended happening that has led, could have led, or could still lead to injury to the patient. These incidents have occurred during or following a medical intervention and are thus iatrogenic: "Induced in a patient by a physician's activity, manner, or therapy"; or induced by health care management in general (e.g., the information systems). Some incidents are the result of a (human) failure in the treatment path: those incidents are referred to as *Medical Errors*. Neither incidents, nor medical errors *all* result in injury. Those that do, are called *Adverse Events*. If an adverse event resulted from a medical error (thus, a subgroup of adverse events), it is called an *Avoidable Adverse Event*. Incidents that did not result in injury are referred to as *Near Misses*. All adverse events are also referred to as *Complications*, but these complications are a wider concept: e.g., also unintended happenings resulting from the patients primary disease or calculated risks are included. The relation between an incident, medical error, adverse event and avoidable adverse event is visualized in Figure 13.

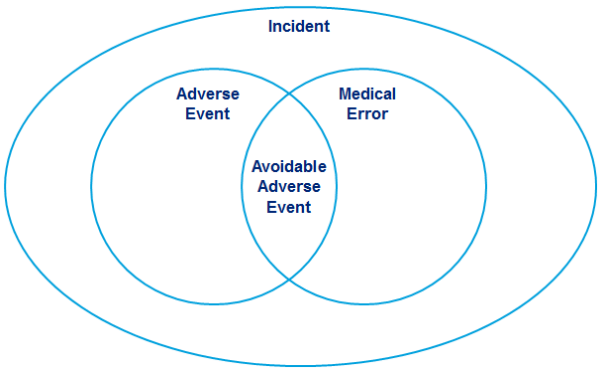

Figure 13: Venn Diagram showing the relationship between the different concepts

The main question of this interview topic is whether the DBC data, which includes information on the operations performed on a patient, provides enough information to identify incidents, and thus patient safety. Can you find incidents from the data, e.g. by identifying an unexpected second operation? How would the data scoe on the contextual quality dimensions *Value Added*, *Relevancy*, *Timeliness*, *Completeness* and *Appropiriate Amount*? What data or information do you structurally miss? Would it be recommended to add these data?

C Used Quotes

The leading language in this thesis is English. However, the expert interviews were held in Dutch. Therefore, the quotes were translated by me. In order not to have any confusions on the interpretation of a certain quote, the mapping between the original quote and the translation are provided in this appendix. Furthermore, a reference to the page of the (200 pages) transcript of all interviews (confidential, only for the thesis supervisors) is provided. With this, the context of the specific quote can be examined more thoroughly, if needed.

The quotes in this appendix are grouped by the expert who stated them. A total of 18 different experts were interviewed. For this thesis, to ensure confidentiality, codings were used to refer to the different experts. The first part of this code is the expert number (e1 to e18), the second part is an abbreviation of the stakeholder group the expert represents (e.g., ”ha” for hospital administration). This appendix is ordered on the expert numbers. The quotes stated by one expert are sorted on appearance in the thesis.

C.1 Expert e1:ha (Tuesday May 22 2012, 14.00-15.30h)

- *”The code in there describing the specialist, that specialist exists <...> there are no strange things in there”*

  Page 182 Die code die daar staat, van de chirurg, die chirurg bestaat <...> *Dus wat dat betreft is het syntactisch niet* nee, zitten er geen gekke dingen in.
- *”The activities released by the care authority, NZA, are very general: it does come close to reality but can be much more detailed.”*

  Page 181 Ik zie er wel problemen in. Die zorgactiviteiten zijn uitgegeven door de zorgautoriteit, NZA, en die zijn dusdanig algemeen dat dat nog wel eens het benadert wel de werkelijkheid, maar dat kan wel verfijnend worden aangepakt
- *”There are also issues with definitions: when can a treatment be called day-care.”*

  Page 183 qua definities is dat bijvoorbeeld ook heel lastig. Een dagverpleging bijvoorbeeld, wanneer is iets een dagverpleging.
- *”De activities are usually quickly registered, within a week I’d say. Opening a DBC is automatically performed, but the coupling of the right diagnosis to the DBC takes a lot of time and effort”*

  Page 183 nou de verrichtingen gaan over het algemeen redelijk snel. <...> dat behoeft wel verbetering maar dat... dat zit er allemaal binnen een week in zou ik zeggen. <...> de dbc is dan geopend maar door de arts moet er dan een diagnose aan gehangen worden en dat kost heel veel inspanning om dat goed voor elkaar te krijgen.
- *”There have been situations where we combined floating activities to DBCs more than a year behind.”*

  Page 184 we hebben nog wel eens zwevende verrichtingen met meer dan een jaar terug zitten te koppelen.
- *”You sometimes notice products are not valid due to the complexity of the treatments in the UMCs. We deliberately decided not complete these products by adding an acceptable activity code. At least, up to now. But this might just change if we are not reimbursed for those treatments anymore. Before <DBC system> it did not matter: we got the money anyway.”*

Page 189 Eh, nou, dat komt voor omdat de dbcs niet compleet gemaakt worden door artsen zeg maar. De dbc ja de validatie niet goedgaat vanwege de ik zal maar zeggen de UMC achtige complexiteit, die je niet kwijt kan. En daarvan hebben wij soms wel echt bewust besloten van ok dan gaan we hem ook niet compleet maken door er iets anders van te maken. <...> tot nu toe wel, maar dat zou zomaar kunnen gaan veranderen als blijkt dat we daar dan geen geld meer voor kunnen krijgen. Kijk voorheen kregen we het geld toch dus maakte het niet uit.

- <About longterm floating activities:> *"If it does not influence the DBC, for example a lab actvitity does nothing, you will not insert them any more. But then these are eventually not in DIS so your profile is not entirely complete"*

  Page 184 Als het verder geen invloed heeft op je dbc, omdat het een labverrichting is dus dat doet verder niets, dan laat je ze gewoon staan. Maar dan heb je ze uiteindelijk niet in je dis zitten. dus je zorgprofiel is dan niet helemaal compleet.

C.2 Expert e2:ha (Monday April 16 2012, 10.30-12.00h)

- *"You want to know something about the DIS. And you want to read things in those data. If you do not realize this <aggregations of registered activities to activity codes> you will think you read things in the DIS that are not there."*

  Page 26 Waarom vertel ik dit, omdat jij natuurlijk iets wil weten over die DIS. En jij wil daar dingen gaan lezen en als jij dit niet beseft, dan denk jij dingen te lezen in de DIS, die er helemaal niet inzitten.

- *I will try to visualize the processes internal to the hospital and the delivery to DIS, then we will automatically get to aspects such as data quality.*

  Page 19 ik denk dat het handig is dat ik probeer hier op het bord de werking intern in het ziekenhuis te visualiseren ook naar die aanlevering DIS en dan kom je vanzelf op punten als datakwaliteit

- *There is a bucket with diagnoses and multiple buckets with the activities: on the OR, policlinic, lab, pharmacy, I can go on for an hour. <...> But you see that the diagnosis, so the label on that DBC, and the activities, are separated from each other in buckets."*

  Page 25 Nou ik heb een bak met diagnoses en ik heb meerdere bakken met verrichtingen, op het OK op het poli op de verpleegafdeling. <...> Op het lab, op de apotheek, zo kan ik nog een uurtje doorgaan. <...> Maar je ziet wel dat die diagnose, dus het etiketje op die DBC en die handelingen, die zitten los van elkaar in bakken.

- *"We perform certain heart surgeries here for which no classification exists <...> you can book these to "nearby codes" "*

  Page 28 we hebben hier hartoperaties die we doen waar nog geen classificaties voor bestaat <...> Je kunt die <onherkenbare verrichtingen> wel boeken naar de nearby codes, zo heet dat dan.

- *"Only everything once declared, is known to DIS. What does not fit the national product structure is not visible. But exists. Only, it does not get further than the validation or grouper phase."*

  Page 27 het enige wat bekend is bij de DIS, is alles wat ooit gedeclareerd is. Met andere woorden, waar in de landelijke productstructuur geen vorm gegeven aan kan worden, zie je dus ook

niet. Maar bestaat wel. Alleen blijft dat hier <validatie of grouper> zitten. Dat komt niet verder dan daar

- *"Billed (Dutch: gefactureerd) is full-continuous the magic word"*

  Page 29 dit plaatje geldt dan ook weer voor de DIS, met dien verstande: gefactureerd. Full continue het toverwoord

- *"I can even approach it minimally: the only activities i register are the typical activities (Dutch: typerende verrichtingen) I need to have a declarable product, I drop the rest"*

  Page 32 Ik kan het zelfs zo minimaal spelen dat ik zeg wat zijn de typerende verrichtingen in de grouper. Het enige wat ik registreer zijn die typerende verrichtingen die ik nodig heb om berhaupt een afleidbaar zorgproduct te hebben de rest laat ik vallen

- *"You should* never *separate the costprice per activity from the corresponding profile. However, this is done in DIS: they take the average national profile of twenty hospitals and couple that with an average cost price per activity of twenty other hospitals"*

  Page 29 Wat is er in de landelijke productstructuur aan weegfout. Die mag je letterlijk opschrijven die krijg je cadeau. Het feit namelijk dat ze het gemiddelde profiel en de gemiddelde kostprijs uit een ziekenhuis hebben losgerukt. <...> Wat wil ik hiermee aantonen, is dat je NOOIT de kostprijs per verrichting mag loskoppelen van het profiel waar die bij hoort. Nochtans, in de DIS, doen ze dat. Want ze pakken daar het landelijk gemiddelde profiel van 20 ziekenhuizen en koppelen die met een uitgemiddelde kostprijs per verrichting van 20 andere ziekenhuizen.

- *"A problem is that it is especially financially focussed. So you are steering to thóse activities that mean something in the DBC game. So less on tangible quality indicators."*

  Page 30 In mijn inleiding heb ik al gezegd die datakwaliteit heeft ontzettende last van het feit dat die zeer financieel getint is. Je gaat dus ontzettend sturen naar die handelingen die leiden in dat DBC spelletje die daar iets in betekenen. En je zit dus veel minder op de minder tastbare kwaliteitsindicatoren.

- *"If you simply think what is in that DIS, it is nothing else than a starting and ending date with a unique typing for the care path, with on the path several activities following an activity code. That is the essence of all that is in there."*

  Page 31 Als je simpelweg full continue denkt wat zit er eigenlijk in die DIS niets anders dan een begin en een einddatum met een unieke typering over dat traject met op dat traject allerlei handeling volgens die zorgactiviteitencode, dan heb ik de essentie vertelt. Dat is wat daar in zit. Niet meer

- *"Then you are getting close to quality issues. Differences in treatment. But you are not going to query those in a reasoned manner. With that I mean you are not going to look at one specific field to find what you are looking for. No you consider the complete data and let it review automatically. Then I give it a chance."*

  Page 32 Dan kom je wel in de buurt van kwaliteits issues. Verschillen in behandeling. Maar die ga je niet op een beredeneerde manier uitvragen. Daar bedoel ik mee je gaat je niet specifiek over 1 veld buigen en kijken ik ga daar vinden wat ik zoek. Nee dan zeg je ik beschouw de totale data en ik ga het vanzelf laten reviewen. Dan geef ik het een kans

C.3 Expert e3:ha (Monday, May 14 2012, 11.00-12.00h)

- *"When it does not conform the syntaxes, it will not even get into <the DIS set>"*
  Page 166 zodra het niet aan die syntaxen voldoet komt het daar niet eens in.
- *"The possible DBC diagnoses are too limited to express the medical complexity <...> it is badly differentiated"*
  Page 167 En de dbc diagnose lijst is redelijk beperkt <...> medisch inhoudelijk zijn die diagnoses beperkt p168: het is slecht gedifferentieerd
- *"The currency depends on the specialism and on the discipline of the specialist. As long as he does not register the diagnosis, it cannot be combined in the DBC."*
  Page 169 het ligt aan het specialisme. Het is ook de discipline van de arts wanneer die registreert. Want zolang die diagnose niet wordt vastgelegd heb ik hem niet in mijn dbc en kan ik hem niet factureren.
- *"Generally, you can pretty much make no statements about this"*
  Page 170 daar kan je in het algemeen redelijk niets over zeggen
- *"Often it is the case that the completeness issues relate to the difficult patients. Those, that do not fit the DBC patterns. These are rejected <by the grouper, Dutch: uitvallen>, but are very interesting"*
  Page 172 eigenlijk kan alleen het ziekenhuis de compleetheid beoordelen. Maar dan zouden ze het terug moeten krijgen. Dan zou je vaker een vergelijk moeten doen, van nou jullie hebben nu 95% van onze gegevens. Het lastige is alleen, dat je weet niet wat die vijf procent is. Zijn dat nou juist de moeilijke patinten. En vaak is dat zo. Die afwijken, die niet in dat patroontje van die dbcs vallen. Dat dat uitvalt. Maar dat wel interessant is.
- *"that is a deficiency in the system: there is 1 diagnosis and the comorbidity of the patient is not in there. <...> The possibility was created for this but I do not think anyone ever delivered one side-diagnosis to the DIS"*
  Page 162 Dat is een beetje een manco in het systeem, je hebt 1 diagnose en zeg maar de comorbiditeit van de patinten zit daar niet in. <...> Ze hebben ooit de mogelijkheid geschept maar volgens mij heeft niemand ooit 1 nevendiagnose aangeleverd.
- *"I'd prefer to see the DBC diagnoses removed and have ICD-10 registrations, although this change is not as straight-forward as it may sound"*
  Page 163 Het liefst zou je willen dat die dbc diagnose eruit gaat en dat je daarvoor in de plaats een ICD-10 registratie hebt. Maar dat heeft wat voeten in de aarde

C.4 Expert e4:ha (Friday, May 11 2012, 10.15-11.30h)

- *"To conclude, after validation not <no syntactic issues>"*
  Page 155 Conclusie, na validatie niet <geen syntactische issues>.
- *"The couplings do not always correspond to the medical reality: there is a risk which has everything to do with the national couplings"*
  Page 155 medisch inhoudelijk kloppen deze koppelingen niet altijd. Daar zit een risico, en dat heeft alles te maken met de landelijke koppelingen
- *"Usually the currency is good. <...> The problem is with activities on ermergency care, which sometimes takes a week before it is put in the system"*

Page 156 Over het algemeen snel. <...> Het probleem zit hem dan met name in verrichtingen die plotseling plaatsvinden, bijvoorbeeld op zon spoedeisende hulp. En daar zie je veel vertragingen. Het kan ook best zijn dat pas een week later pas de administratie wordt doorgenomen <...> , dan wordt het pas een week later in het systeem gezet

- *"Required fields are filled, although there might of course be semantic problems. <...> The facultative fields are, with high probability, not filled."*

  Page 158 Verplichte velden zijn ook verplichte velden in het hele registratieproces. Dus die moeten gewoon ingevuld worden. <...> <hoe zit het met de onverplichte velden?> Eh daar is de kans groot dat het niet wordt ingevuld.

- *"This also relates to the academic culture: if something is not mandatory, specialists will rarely put effort to register it"*

  Page 158 Dat heeft met name ook weer met de academische cultuur te maken. Als iets niet verplicht is, zullen artsen zelden tot niet de moeite nemen om het dan toch in te vullen

- *"Possibly, DIS should have more fields required. If you want to prepare for mandatory ICD-10 registrations, you have to require that ICD-10 field already this year or the hospital will not start working on the delivery of it."*

  Page 158 misschien moet DIS wel meer verplichten. He want inderdaad als je je wil voorbereiden op verplichte registratie ICD-10, dan moet je in 2012 nu al dat veld ICD-10 of de diagnose verplichten.

- *"I think this is the biggest issue, it is not good. Sometimes things are forgotten, especially activities, or they are not placed in the right trajectory or disappear in the system and stay only in the internal database. All that, is not submitted to DIS"*

  Page 158 Dit is eigenlijk denk ik het grootste, de grootste onvoldoende laat ik het zo zeggen. Dit is namelijk niet op orde, dus laten we zeggen vier. Er wordt nog wel eens wat vergeten, met name verrichtingen, of verrichtingen komen niet in de juiste winkelwagen terecht, die verdwijnen in je systeem, blijven in de database achter. En dat alles wordt niet meegenomen richting DIS.

C.5 Expert e5:ha (Monday, April 23 2012, 10.00-12.15h)

- *"The whole healthcare sector is combined with tables, you see that in the datamodel"*

  Page 65 de hele zorg hangt met tabellen aan elkaar dat zie je ook al aan dat hele datamodel enzo

- *"It is all validated and an activity code should be from a table. <...> There is almost no free text in there, it can hardly go wrong."*

  Page 65 <...>dat wordt helemaal gevalideerd en een verrichtingscode moet uit een tabel komen. Dat is eigenlijk allemaal redelijk gestandaardiseerd. Daar zit bijna geen vrije tekst in, dat kan bijna niet misgaan zou je zeggen.

- *"There are many more risks in that dimension <Semantic Accuracy>"*

  Page 66 Kijk als ik verwijzers moet vastleggen en ik heb twee of drie dokters Jansen in Den Haag en ik kies de verkeerde, dan heb ik een verkeerde verwijzer bij een DBC. Syntactisch klopt het natuurlijk. Het is een huisarts, hij staat in de tabel, hij heet Jansen, alleen hij woont ergens anders. Ja daar heb ik niet naar gevraagd bij de patint. Daar <semantische accuraatheid>

bestaan veel meer risicos.

- *"It's a financial system, the output are Euro's and not to determine whether a patient has been improved. It is a translation of the actions performed by the specialist with as goal number one to send an invoice."*

  Page 64 Het is een financieel systeem, de output zijn euros, de output van deze registraties zijn niet om te kijken of de patint beter is geworden. Het is een vertaalslag van wat de dokter doet met als doelstelling nummer 1 er moet een factuur de deur uit.

- *"Usually, there are sub systems. For example, a lab system. A labrequest is performed and executed, but it can take up to a week before it is in the system. I think that is too long, it should be one day (OLTP system)."*

  Page 70 over het algemeen hebben we te maken met subsystemen, zon labsysteem bijvoorbeeld, daar wordt een labaanvraag gedaan, die wordt dezelfde dag uitgevoerd. En dan hebben we de situatie dat hier in huis bijvoorbeeld duurt het een week voordat die labverrichting in het systeem zit. Dat duurt me allemaal veel te lang, want eigenlijk wil je alles wat je vandaag doet, dat morgen in het systeem zit. Het is een transactiesysteem, dus daar gaan allemaal nachtverwerkingen overheen, dus je kan niet sneller dan morgen die data bij elkaar hebben

- *"The required fields are validated so should be in. <...> But you have to be careful when you want to utilize facultative fields in a research"*

  Page 73 Die verplichte velden worden gevalideerd, dus die moeten erin zitten <...> Betekent ook dat als je met onderzoeken gaat werken met niet verplichte velden, dat je goed moet uitkijken wat je aan het doen bent

- *"Eventually you will miss a couple of percentages. On rejections, or corrections from insurers that are not submitted"*

  Page 74 Uiteindelijk ga je natuurlijk een paar procent missen. Aan uitval, en aan niet doorgevoerde correcties, en aan niet meer aangeleverde correcties.

- *"This <missing activities, e.g. due to rejections> might be the case, but if you look at relative missing entities it is only one or two percent. Because eventually something is made up for these activities anyway."*

  Page 74 En dat is ook wel zo, en dat maakt misschien dat je die data een 8 moet geven in plaats van een 10, maar ga je kijken naar de relatieve missende regels ja dan praat je misschien over een procent, een missende twee procent. Omdat uiteindelijk, mensen toch wel wat van uitval weten te maken.

- *"You end up with like 98,5% of the data that passes all validations and which is used for the financing of the hospitals. Then, you can hardly keep saying that the data is hopeless. I do not believe that"*

  Page 74 dan heb je best een hoop data, 98,5 procent, dat door alle validaties heenkomt wat voor de financiering van ziekenhuizen wordt gebruikt. Dan kan je moeilijk volhouden, dat is hopeloze data. Dat gaat er bij mij niet in.

- *"You should actually require in the NZA model for cost prices that* all *activities in the DBC Maintenance activity table determine the costs. Instead of not registering half of the activities and attaching their costs to the other half. Which is what happens."*

  Page 76 Dat is natuurlijk wat je in het NZA model van kostprijzen keihard moet vereisen dat je

verrichtingen die in de verrichtingen tabel van DBC Onderhoud staan dat je die als kostendragers mee gaat nemen. En niet gaat zeggen van de helft doe ik maar niet en de kosten daarvan sleep ik over naar de andere helft. Dat is wel wat gebeurt natuurlijk

- *"with every care request, the "requester" is provided, which is directly registered. There is no control whether this requester has an existing DBC. That is the problem. The whole workflow management system, the communication of orders, is absent in most hospitals. Therefore, everything is based on a transactional system of registering a care request together with the requester. This problem is everywhere and one of the largest causes of floating DBCs"*

  Page 70 Bij elke aanvraag wordt aangegeven wie de aanvrager is. Die wordt opgenomen. En dan is er geen controle op van heeft die wel een DBC. Dat is het probleem. Het hele workflow management systeem zeg maar, de ordercommunicatie, dat is in heel veel ziekenhuizen nog niet aanwezig, dus alles gaat nog op basis van transactiesysteem van het invoeren van een aanvraag, inclusief de aanvrager. 50.00 Dit probleem speelt overal. Is een van de grootste oorzaken van zwevende DBCs.

- *"The question is if that, due to the age, leads to a change in the DIS delivery. <...> You would want to see that the hospital deliveries to DIS also contain corrections of older years. Because it would be really strange if a hospital never submits something older than two years. All would have been submitted good at once, and trust me, that is not the case"*

  Page 68 Maar de vraag is maar of dat gezien de ouderdom leidttot een verandering in de DIS aanlevering <...> je zou ook willen zien dat bij die ziekenhuisaanleveringen aan DIS ook correcties van oudere jaren zitten. Want het zou toch heel gek zijn als er ziekenhuizen zijn die nooit wat aanlevert ouder dan twee jaar. Dat zou dan allemaal in n keer goed gaan. En geloof me, dat is niet zo.

- *"These data should be in the DBC system much faster. By increasing the communication speed between the different systems. By having more tight procedures like always registering the agenda contents at the end of a consultation hour. Instead of saying we wait a week with this, because there are always corrections and such. That is only because the process is not streamlined."*

  Page 71 Door systemen veel sneller op elkaar aan te laten sluiten, procedures veel strakker insteken dat je zegt van ja een agenda afwerken moet je gewoon op het eind van een spreekuur doen, dat het de volgende avond gewoon in het dbc systeem staat. In plaats van dat je zegt van nou ja goed we wachten er een week mee. Omdat je toch allemaal correcties hebt en dergelijke. Dat komt gewoon omdat je proces niet op orde is.

- *"DBC Maintenance currently receives signals of all kind of things that due to issues and innovations are not deduced correctly by the grouper. With this, they start working on it with different stakeholders. But they never do that based on data of rejections (Dutch:* uitval*). It is always based on incidents. While you could also say like, we are busy three months and have three months of data, we will analyze everything and see what goes wrong"*

  Page 72 Kijk wat nu er nu gebeurt is dat DBC Onderhoud door middel van allerlei issues, innovaties, signalen krijgt van allemaal dingen dat niet lekker door die grouper heenkomen. En gaan daarmee aan de slag met partijen. Maar zij doen dat nooit op basis van data van zaken die uitvallen. Het gaat altijd op incidenten van dingen die gebeuren. Terwijl je ook kan zeggen van weet je wat, we zijn drie maanden bezig, we hebben alle data van die drie maanden binnen, we gaan analyseren van wat gaat er fout

- *"So actually, DIS should say, we do not mind we want everything. Even DBCs that are not closed, also DBCs that were not deduced to a valid product, even the floating activities."*
  Page 72 Dus eigenlijk zou DIS moeten zeggen we vinden het niet erg, we willen gewoon alles. Ook DBCs die niet gesloten zijn, ook DBCs die niet tot een valide product hebben geleidt, ook zwevende verrichtingen.
- *"You would want to state: I expect at least these types of activities, approximately in those amounts. This is a validation you could do before performing analyses. In all research you perform on these data, you will experience that you need to check for completeness."*
  Page 64 Je zou willen stellen: ik verwacht minstens dit soort type verrichtingen, ongeveer in die aantallen. Die validatieslag zou je van tevoren kunnen doen. Alle onderzoek die je doet met deze DIS data, daar zal je tegenaan lopen dat je moet controleren op volledigheid.

C.6 Expert e7:bi (Friday, April 27 2012, 10.15-12.30h)

- *"When extracting the data some things might go wrong but these are rejected: <syntactically> it is of good quality"*
  Page 106 Bij extraheren van data kan er wel eens wat fout gaan maar dat krijg je dan ook terug. Kwalitatief zit het goed in elkaar.
- *"The last two years you can better assume the activity code represents the actual performed activity"*
  Page 106 nu <de laatste twee jaar> kun je er een stuk beter vanuit gaan dat de verrichtingscode dat er staat, ook daadwerkelijk weergeeft wat er is gedaan
- *"Although semantic accuracy has increase, there are still questionmarks related to it. These points and examples will be used by hospitals to undermine analyses on these data."*
  Page 106 <semantische> accuraatheid wel toegenomen, maar er staan nog wel vraagtekens bij. Dit soort punten/voorbeelden zouden door ziekenhuizen al snel aangehaald worden om de kwaliteit van de gegevens van een externe partij (die de gegevens heeft geanalyseerd) onderuit te halen. (Although the last two years)
- *"Less nuances are needed in general hospitals to represent your activities <...> We (UMC) often do innovative operations that do not suit any existing ZA code"*
  Page 106 in periferie heb je veel minder nuances nodig om aan te geven wat je aan het doen bent <...> P106 In het UMCG doet men veelal iets dat nog geen ZA codering heeft: kan men niet kwijt. A doordat academisch ziekenhuis is wat nieuwe dingen doet die je nog niet ergens onder kan brengen, wat trouwens ook lastig is voor de financiering.
- *"It is possible that certain activities are well combined to a DBC, where a combination to another DBC would be preferred as it would result to a better care product"*
  Page 107 de kwaliteit van de koppeling: Het kan heel goed dat allerlei verrichtingen heel goed koppelbaar waren aan een DBC, maar dat je liever had gehad dat dat aan een andere DBC zou zijn gehangen, zodat er dan een beter zorgproduct uit zou zijn gekomen.
- *"We submit all required fields. Regarding the facultative fields, we only submit those items we can easily generate"*
  Page 110 leveren alles aan wat moet. Dat krijgen ze sowieso. <...> Van de niet verplichte items leveren ze alleen datgene aan wat al ze standaard kunnen ophoesten.

- *"Patients and DBCs are very good (all exist in DIS). But if you look at the combinations of patient-carepath-subtrajectory-activity: everything is in what should be in. But you do miss data especially the very specific activities such as those from neurosurgery. <...> This does not occur often but these are important, e.g. activities with which we can justify our academic setting"*

  Page 110 maar als je kijkt naar patint-zorgpad-dbc-verrichting: Ja, alles staat erin wat erin moet staan. <...> je mist wel data, vooral voor die hele specifieke dingen van neurochirurgie en van oncologie ook een paar dingen. <...> <MAIL:> Dit komt voor maar in beperkte mate vervelend hierbij is dat dit soms wel erg essentiele verrichtingen zijn waarmee wij bijvoorbeeld ons academische setting kunnen verantwoorden. Aantalllen zijn niet groot maar wel erg belangrijk

C.7 Expert e8:bi (Wednesday, May 2 2012, 13.15-14.15h)

- profile: *"How many activities are on average performed for a specific DBC"*

  Page 118 zorgprofielen: hoeveel verrichtingen doe je nu gemiddeld voor een DBC.

C.8 Expert e9:dis (Wednesday, April 18 2012, 10.15-11.30h)

- *"The invalid <data rows> are not stored"*

  Page 43 Hetgene wat fout is, slaan we niet op.
- *"Some of the fields are required: these are certainly all in the DIS set, this is validated before entry. However, also non-required fields exist"*

  Page 48 er zijn verplichte velden, dat is sowieso compleet daar wordt ook op gevalideerd. Maar er zijn ook niet verplichte velden, zoals declaratiedatum
- *"For us, it is very interesting that the hospital says this. Because, for us it is not rejected (Dutch:* uitval*) because it's not in our scope up to this moment. But it would be nice if it would become in our scope, so we could recognise and call it as* uitval*."*

  Page 40 voor ons is het wel heel interessant dat het ziekenhuis dat zegt juist. Want voor ons is het geen uitval omdat het tot nu toe niet in de scope is om zo te zeggen. Maar het zou mooi zijn als het wel binnen de scope komt. Want dan kunnen we het als uitval gaan benoemen.

C.9 Expert e10:dis (Wednesday, April 18 2012, 10.15-11.30h)

- *"we know certain DBCs are clinical treatments so we expect treatment days. Or one of those activities. If that is not the case, we put a flag there"*

  Page 37 Dan wordt er echt heel inhoudelijk naar de data gekeken. Dat wil zeggen, we weten dat bepaalde dbcs onder bijvoorbeeld klinische opnames vallen, dan verwachten we ook dat daar verpleegdagen bij zit. Of 1 van de verrichtingen. Als het niet zo is, dan zetten we daar een vlaggetje bij.
- *"<there is validation> whether a code that is being used, an activity code: it should be an existing activity code. <...> Only when this is all OK, it is stored."*

  Page 43 ja, <validatie dat> of er een code wordt gebruikt, een activiteitencode, dan moet dat een bestaande activiteitencode zijn in de codelijst van DBC Onderhoud. Peter: dus die

kwaliteitscheck doen jullie, intern [ja] bij die validaties [ja].Dick: ja als dat allemaal OK is, dan pas wordt het doorgeladen en wordt het bij ons opgeslagen.

- *"These kind of mistakes, intrinsic to the DBC system, are not easy to identify and dissolve"*

  Page 47 Als je praat over inhoudelijke fouten, dan heb je het over de DBC systematiek, is het <zoeken van fouten> niet bepaald eenvoudig. En dus is het ook niet eenvoudig om die fouten eruit te halen.

- *"If you are talking about the completeness of the data you can differentiate two levels: DBCs and activities. On DBC level we will score pretty high. This has externally been assessed once and it was a good match. <...> On detail level, really the activities <...> you can't really say something about it, it is differentiated and depends on how well a hospital has organized his own administration"*

  Page 47 als je praat over de volledigheid van gegevens, dat kan je eigenlijk ook wee ronderscheiden naar dbc niveau en activiteit niveau. Dan is op dbc niveau scoren we denk ik ook vrij hoog. Een extern bureau heeft dat wel eens assessed, en dat kwam redelijk overeen. <...> Voor detailniveau, dus echt verrichtingen: <...>daar kan je natuurlijk helemaal niets van roepen, of het wel of niet volledig is. <...> dat is gedifferentieerd, ligt er aan "hoe goed houdt een zorginstelling zijn eigen administratie op orde".

- *"This is mainly the case in university hospitals. Those perform relatively more activities that are not defined, so these have more rejections"*

  Page 40 Dat komt vooral bij academische ziekenhuizen voor. Want die doen vaak naar verhouding veel meer verrichtingen die een beetje buiten de boot vallen. Dus die hebben veel uitval zeg maar

- *"The second pseudonymization key is typical to DIS and always the same. The CBS also has pseudonymized data with a CBS pseudonym. If we deliver to the CBS, it goes through ZorgTTP. These remove the DIS pseudonym and place the CBS pseudonym. Through this for example the social security number (SSN), although still pseudonymized and thus not directly recognizable, can be coupled to the SSN pseudonym the CBS already has."*

  Page 39 het tweede pseudoniem is typisch DIS, dus al de data die naar dis gaat. De tweede sleutel die eroverheen gaat die is voor DIS eigenlijk altijd hetzelfde. Maar het CBS heeft ook gepseudonimiseerde data met een CBS pseudoniem. Als wij dus terugleveren aan het CBS gaat dat weer via zorgttp, die halen het DIS pseudoniem eraf, zetten het CBS pseudoniem er weer op, waardoor bijvoorbeeld het BSN nummer, het pseudoniem dat dat is, dus niet direct herkenbaar, maar wel koppelbaar is met het BSN pseudoniem wat het CBS al heeft.

- *"We never see this! It is an observation which can only be made by a hospital, not by DIS."*

  Page 40 wij zien dit natuurlijk nooit he. Het is alleen maar een constatering die een ziekenhuis kan doen, en dat kan DIS niet doen.

- *"Actually, we once offered them to send all those activities, to see if we could do something with it. It is easy to say that one has so much or so few, or how much underivable products (Dutch:* uitval*) there are. But you are only sure once you start measuring."*

  Page 40 sterker we hebben wel eens de aanbieding gedaan van stuur dat maar op naar ons. Al die verrichtingen die jullie dan kijken we wel of we daar iets mee kunnen doen. Het is natuurlijk makkelijk te zeggen van ja we hebben zoveel of zo weinig, of hoeveel uitval er is. Maar dat weet je pas zeker als je het gaat meten. Meten is weten.

- *"The problem is mainly with the delivered activities. One DBC should contain multiple activities. We know that this is incomplete for certain hospitals and not as it should be."*

  Page 42 Maar waar de [voet daar nog in klem zit? - Minute 23 of recording] is de meegeleverde zorgactiviteiten. In een DBC. Dus aan een dbc hangen 1 of meerdere activiteiten als het goed is. En daar weten we van dat het bij sommige ziekenhuizen incompleet is en niet zoals het zou moeten zijn.

C.10 Expert e11:dis (Wednesday, April 18 2012, 10.15-11.30h and 11.30-12.30h)

- *"There are no specific codes whether an incident has occurred. This is not relevant for the reimbursement."*

  Page 54 maar er wordt niet zozeer geregistreerd of er echt een incident is geweest, daar hebben wij gewoon geen codes voor zeg maar. Dat maakt voor de financiering ook niet uit voor de betaling.
- *"It <DIS> gives you a lot of information but I doubt whether it is relevant for safety management. Not as it is. <...> It is financial information, not directly useful."*

  Page 60 het geeft je een hele hoop andere informatie maar of dat relevant is voor het veiligheidsmanagement, dat vraag ik me af. Niet zoals het er nu instaat. <...> Het is financile informatie en daar kan je niet direct wat mee.
- *"The patient is recovered, but the question is whether an extra activity is registered for that"*

  Page 15 De patint wordt herstelt, het is maar de vraag of daar een extra activiteit voor wordt geregistreerd
- *"Do you then have a more expensive product because you have a more complicated patient, or because you have dropped him? <...> You will really have to inspect what happened in the individual cases before you can draw conclusions"*

  Page 59 Is het dan een duurder product omdat jouw patint ingewikkelder is, of heb je een duurder product omdat je die patint een keertje hebt laten stuiteren? <...> Page 55: Dan zul je toch echt per geval moeten bekijken wat is er gebeurd voordat je conclusies kan trekken.

C.11 Expert e12:hc (Friday, April 13 2012, 10.00-11.30h)

- *"It is a hospital administration goal. Which is, by definition, different than <eh> quality or safety as a goal. These systems are not developed to extract quality or safety indicators; but distinct systems that are not administratively related."*

  Page 8 het is echt een hospital administration doel. En dat is per definitie anders dan eh.. kwaliteit of veiligheid , als doel. <...> Maar die systemen zijn niet gemaakt om eh veiligheids of andere kwaliteits indicatoren eruit te halen. <...> Daarvoor zijn aparte, niet administratief gerelateerde, systemen.
- *"A lot of this patient safety information, apart from volumes, can not be derived from the basic health administration, as they are stored in other separate systems. Partly because those systems, with regard to safety very consciously, are not freely accessible."*

  Page 13 Een hoop van de patintveiligheidsinformatie, los van volumes en revisietermijnen, is op dit omment nog niet te halen uit de standaard bedrijfsinformatie, omdat ze in andersoortige

(separate) systemen worden bijgehouden. Deels ook omdat die systemen, bij veiligheid juist heel bewust, niet toegankelijk zijn.

C.12 Expert e13:hc (Wednesday, May 9 2012, 13.00-14.30h)

- *"This implies that when something goes wrong or almost wrong on a department, which can be medical but also in the laboratory or radiology, it is reported in that specific stand-alone system"*

  Page 141 Dat betekent dat als er op een afdeling iets mis gaat of bijna iets mis, dat kan medisch zijn, verpleegkundig, maar ook in het laboratorium of een rntgen of waar dan ook, dan wordt dat gemeld in dát systeem. Dat is een apart systeem, een stand-alone systeem
- *"You want a specific culture in the hospitals to allow the employees to report incidents safely. If you ruin that culture <...> he might consider the next time not to report everything any more".*

  Page 142 Maar je wil ook een bepaalde cultuur in het ziekenhuis hebben dat de medewerkers veilig kunnen melden. Als je die cultuur verpest <...> die denkt de volgende keer misschien van ik ga niet alles meer melden
- *"I do not think it is in there. At least not in this hospital, it is really a separate system. Everybody knows that here, incidents are reported in the risk management system, this is for the health care administration."*

  Page 142-143 Ik denk niet dat dat daar inzit. Niet in dit ziekenhuis althans want dat is echt standalone, dat is een apart systeem. Dat weet ook iedereen, incidenten worden gemeld in het risicomanagement systeem en dit is echt voor de zorgadministratie.
- *"It is also possible the patient has fallen out of bed the second day. Or received the wrong nutrition. Or has called three times at night while the nurse did not show up... you will never be able to get those kind of incidents from this administration. It will require a more in-depth analysis in other administration, you will really have to perform analysis on patient records."*

  Page 143 Maar het kan ook dat een patint bij ons de tweede dag uit bed is gevallen. Of hij heeft verkeerde voeding gekregen. Of hij heeft s nachts 3x gebeld en de verpleegkundige kwam alsmaar niet. . Dat soort incidenten zul je nooit uit deze administratie kunnen halen. Zul je altijd de diepte in moeten, in andere administraties. Dan moet je echt een dossieronderzoek gaan doen.
- *"You will not find in these data when a leg has been amputed from the wrong patient: we will do that on own accounts instead of sending an invoice to the insurer"*

  Page 147 Je zal in deze data ook niet terug kunnen vinden wanneer bij een verkeerde patint een been is geamputeerd: dat zullen we dan wel op eigen rekening doen in plaats van bij een zorgverzekeraar die beenamputatie te declareren.
- *"Yet another example. This is often what the women chose themselves"*

  Page 144 (over eerst borstbesparende dan borstamputerende operatie) Daar noem je er ook weer 1. Dat is ook vaak wat die vrouwen zelf willen.
- *"In those cases you could ask yourself "what has happened here"? But then you would probably still need to analyze separate systems to determine what exactly happened"*

  Page 143 <Zou je kunnen denken> Wat is hier gebeurt? Maar dan is nog wel de vraag of je dat uit deze informatie kan halen om te onderzoeken wat er dan is gebeurd. Daar zul je dan waarschijnlijk ook nog separate systemen voor moeten raadplegen.

- *"You could indicate which care paths or DBCs have an increased chance for incidents based on these data, but that is too short-sighted. It could be caused by so many things, you would have to look deeper into the data and these data do not provide that possibility"*

  Page 147 Je zou aan de hand van de data misschien wel kunnen aangeven welke zorgtrajecten of DBCs verhoogde kans op een incident hebben, maar dat is veel te kort door de bocht. Dat kan aan zoveel dingen liggen, je moet vervolgens dieper die data in kunnen kijken en die mogelijkheid biedt deze data niet.

C.13 Expert e15:con (Thursday, April 12 2012, 12.00-13.00h)

- *"Take into account that it has been developed from a financial point of view. DBC, and DOT, is a financial system, developed from the need for a new reimbursement system."*

  Page 6 Hou er rekening mee dat het is ontwikkeld vanuit een financieel oogpunt. DBC, en DOT, is een financieringssysteem, ontwikkeld vanuit de behoefte aan een nieuwe financieringssystematiek.
- *"You do not know the cause of say extra hospitalization. For example, are those caused by something irrelated such as an eighty year old choking in some food and getting in shock, or because of bedsores after the patient wasn't cleaned for one day."*

  Page 6 Je weet niet de reden van bijvoorbeeld extra ligdagen. Bijvoorbeeld, zijn die extra ligdagen veroorzaakt door een losstaand iets als een shock ten gevolge van een 80 jarige die zich verslikt, of omdat er een doorligplek (incident) is ontstaan nadat hij een dag niet is verschoond.

C.14 Expert e16:sci (Tuesday, May 8 2012, 15.15-16.15h)

- The second approach, Free Disposable Hull, is more restricted. In this approach *"You only qualify a hospital better than another one if it delivers (at least) the same production but uses less of* all *inputs"*

  Page 133 Free Disposable Hull <...> daar zitten dus helemaal aan de conservatieve kant met je efficiency schattingen. Want dan zeg je van ik vind een ziekenhuis of een instelling alleen beter dan de ander als die minimaal dezelfde productie levert en van alle inputs iets minder gebruikt
- *"I do not see anything at all that I could use which has something to do with inputs"*

  Page 130 Ik zie helemaal niets daar waar ik iets aan zou kunnen hebben wat iets met die inputkant te maken heeft.

C.15 Expert e17:sci (Monday, May 22 2012, 14.00-15.00h)

- *"I see the DIS data mainly as an output measurement. <...> Maybe activities as input.*

  Page 177 ...die DIS data zie ik eigenlijk vooral als de output maatstaf. <...> Ik zit te denken van <...> inputs van verrichtingen.

C.16 Expert e18:sci (Tuesday, April 24 2012, 13.00-14.00h)

- *"So it should have resulted in a longer hospitalization, physical injury, a wrongly amputated leg, that category"*

  Page 85 Dus het moet daadwerkelijk, langere ligduur, fysieke schade, een verkeer arm of been afgezet dat idee
- *"On average, it takes a nurse half an hour to analyze a record, and possibly another half an hour extra for the analysis by the specialist."*

  Page 88 Als je globaal kijkt is het ongeveer een half uur voor een verpleegkundige voor een dossier, en een half uur voor een specialist, er eventueel bij.
- *"You will probably not recognize in the DBC data whether a patient has filed a complaint. Those are typically aspects you get from the patient records."*

  Page 86 Wat je waarschijnlijk niet in de DBCs terug zal vinden is of de patint een klacht heeft geuit. Nou dat zijn echt dingen die je dan echt weer uit dossiers haalt.
- *"But you never see in the DBCs, and not in the patient record either, what the background was, what were the causes"*

  Page 97 Maar je ziet nooit in de DBCs en ook niet in de dossiers wat nou de achtergrond was uiteindelijk, de oorzaken
- *"If you really want to see whether damage occurred, you will have to look into the specific record"*

  Page 89 Wil je echt kijken van is er daadwerkelijk schade geweest, dan moet je het een op een terugkoppelen naar het dossier
- *"it might not imply that something has certainly gone wrong, but it does stand out <...> This is also what the IGZ uses, they say that stands out, show us your detailed data. Then they start an inspection"*

  Page 89 Het wil niet zeggen dt er wat mis is gegaan, maar het valt wel op. Page 95: Zo gebruikt de IGZ dat ook, die zegt van dat valt op, laat je gegevens maar zien. Dan

D Care Profile Classes

Below is a complete overview of the available care profile class codes (Dutch: *ZorgProfielKlasses, ZPK codes*) in the Dutch DBC and DOT systems.

ZPK	Definition	(Dutch)
1	Outpatient department / ER	Polikliniek- en eerste hulpbezoek
2	Daycare	Dagverpleging
3	Clinic	Kliniek
4	Diagnostic activities	Diagnostische activiteiten
5	Surgical activities	Operatieve verrichtingen
6	Other therapeutic activities	Overige therapeutische activiteiten
7	Medical imaging	Beeldvormende diagnostiek
8	Clinical chemistry and hematology	Klinische chemie en haematologie
9	Microbiology and parasitology	Microbiologie en parasitologie
10	Pathologie	Pathologie
11	Other laboratory operations	Overige laboratoriumverrichtingen
12	(Para)medical functions	(Para)Medische en ondersteunende functies
13	Prosthetic implants	Bijzondere kunst- en hulpmiddelen
14	Rehabilitation	Revalidatie
15	Blood products	Bloedprodukten
17	Long Asthma Centers	Longastmacentra
18	Other ER activities	IC zorgactiviteiten niet zijnde ic-behandeldag
19	ER treatment	IC-behandeldag
20	Expensive drugs/medicine	Dure geneesmiddelen
21	Other drugs/medicine	Weesgeneesmiddelen
22	Clotting factors	Stollingsfactoren
89	Other activities	Overige zorgactiviteiten t.b.v. afleiding
99	Not included in careprofile	Niet in profiel meegenomen

Table 20: ZPK code overview (adopted with permission from [86])

E Incident Triggers

#	Trigger
1	Patint was reeds eerder (¡12 maanden) opgenomen omreden(en) gerelateerd aan de indexopname
2	(Aanvankelijk) onbedoelde opname binnen 12 maanden na ontslag van de indexopname
3	Blijvende of tijdelijke schade opgelopen tijdens het index ziekenhuisverblijf
4	Een schadelijke en onbedoelde reactie op een geneesmiddel
5	(Aanvankelijk) onbedoelde overplaatsing van een algemene afdeling naar de intensive care (inclusiefhartbewaking)
6	(Aanvankelijk) onbedoelde overplaatsing naar een ander ziekenhuis na een onverwachte verslechtering van de patiënt
7	(Aanvankelijk) onbedoelde (her-)operatie
8	(Aanvankelijk) onbedoelde verwijdering, beschadiging of herstel van een orgaan of weefsel tijdens een operatie of invasieve handeling
9	Infectie/sepsis tijdens het index ziekenhuisverblijf (exclusief infectie/sepsis opgetreden binnen 72 uur na opname)
10	(Andere) complicaties, bijvoorbeeld acuut hartinfarct, TIA/CVA, longembolie enz. (behelst elke onverwachte complicatie die niet een natuurlijk gevolg is van de ziekte van de patint of een verwachte uitkomst van de behandeling)
11	Neurologische afwijking ontstaan tijdens indexopname (behelst neurologische afwijkingen
12	(Aanvankelijk) onverwacht overlijden (geen sprake van opname voor palliatieve zorg)
13	Hart-/ademstilstand (en reanimatie succesvol)
14	Niet passend ontslag naar huis/inadequate ontslagplanning bij de indexopname (medisch onverantwoord ontslag; exclusief tegen medisch advies in)
15	Ontevredenheid over de zorg gedocumenteerd in de dossiers en/of aanwijzingen voor ingediende klachten (inclusief gedocumenteerde klacht, conflicten tussen patint/familie en staf, ontslag tegen medisch advies in)
16	Alle andere ongewenste uitkomsten die hierboven niet worden genoemd

Table 21: Triggers pointing to care related damage as used in [56]