

TRANSFoRm

Translational Research and Patient Safety in Europe

The TRANSFoRm Project is partially funded by the European Commission under the 7th Framework Programme
Grant Agreement Number FP7-247787

D5.1

Data Quality Tool

Work Package Number:	WP5
Work Package Title:	INTERFACES: USER AND SOFTWARE SERVICES (SO5)
Nature of Deliverable:	Report
Dissemination Level:	Confidential
Version:	1.1
Delivery Date From Annex 1:	M36
Principal Authors:	N.A. Khan, M.McGilchrist, S.Padmanabhan T.van Staa, R.A. Verheij,
Contributing Authors:	A. Algharibi, T.N. Arvanitis (University of Birmingham),
Partner Institutions:	NIVEL, University of Dundee, CPRD,



7th Framework Programme <http://cordis.europa.eu/fp7/ict/>
European Commission http://ec.europa.eu/information_society/



Contents

List of Figures	4
List of Tables	5
Executive summary	7
1 Introduction and structure of the report	8
2 Purposes of EHR data and factors influencing data quality	9
2.1 Purposes of using primary care EHR data	9
2.2 Factors influencing data quality: a stepwise approach.....	10
2.3 Conclusion	15
3 Data quality framework	15
3.1 Introduction	15
3.2 The role of scientific or managerial context	18
3.3 Axiomatic basis – purpose and population	19
3.4 Population determination	20
3.5 Measurement Value Set (MVS).....	21
3.6 Characterisation of a measurement value set.....	22
3.7 Metrics.....	24
3.8 Visualising the metrics	27
3.9 Contextual provenance of data	29
3.10 Data quality statements	31
3.11 Tasks for determining data quality.....	33
4 Determining Data Quality for TRANSFoRm Diabetes Use Case.....	35
4.1 Databases	35
4.2 Purpose and population.....	36
4.3 Measurement Value Set (MVS).....	38
4.4 Metrics.....	39
4.5 Visualising the Quality metrics and Quality Statements	40
4.6 Discussion	46
5 TRANSFoRm Data Quality Webtool.....	47
5.1 Requirements for the Data Quality Webtool.....	47
5.2 Architecture of the Data Quality Webtool.....	48
5.3 User Interface Design and Workflow of the Data Quality Webtool	50
5.4 Evaluation of the Data Quality Webtool.....	55
Appendix A: NIVEL primary care database: classifications and coding	64

Appendix B: GPRD database: coding and classifications 66

Appendix C: TRANSFoRm use case GORD..... 67

 Introduction 67

 Purpose and population 67

 Measurement Value Set (MVS) 68

 Metrics..... 68

 Visualising the Quality metrics and Quality Statements..... 69

Appendix D: TAM-inspired questionnaire for the Data Quality Tool

Evaluation 72

References..... 77

List of Abbreviations 80

List of Figures

FIGURE 1 THE OVERALL PROCESS FROM PATIENT ACCESS AND MEASUREMENT IN A CLINICAL SETTING S(HCP) TO END USE BY THE RESEARCHER (USE). CZ – CARE ZONE; NCZ – NON-CARE ZONE, OR LINKAGE ZONE; RZ – RESEARCH ZONE. S(HCP) – HEALTH CARE PROFESSIONAL; S(GPIS) – GP INFORMATION SYSTEM; S(REPOS) – CLINICAL DATA REPOSITORY SYSTEM; S(TRANS) – TRANSFORM SYSTEM; USE – END USER.....	11
FIGURE 2 ILLUSTRATION OF THE RELATIONSHIP BETWEEN THE CORE CONCEPTS OF THE DATA QUALITY MODEL. A1 – AXIOM 1, A2 – AXIOM 2.....	20
FIGURE 3 THE PROPORTION OF TYPE-2 DIABETIC PATIENTS WITHIN A GIVEN PRACTICE HAVING AT LEAST ONE MEASUREMENT OF A GIVEN TYPE (HERE HBA1C) WITHIN A GIVEN YEAR (2009) IS COMPUTED AS A MEASURE OF COMPLETENESS. QUALITY DETERMINANTS IN THIS FIGURE INCLUDE PRACTICE AND EHR VENDOR. TAKEN FROM THE NPCD DATABASE OF NIVEL.....	28
FIGURE 4 BOX PLOTS VISUALISING THE DISTRIBUTION OF COMPLETENESS FOR VARIABLES RELEVANT FOR THE DIABETES USE CASE.....	29
FIGURE 5. THE DATA QUALITY MODEL.....	33
FIGURE 6 PERCENTAGE OF NON-TEMPORARY PATIENTS AGED 18+ WITH DIABETES CODE (EVER IN THEIR RECORD) IN 2009, 2010, AND 2011 IN THE NPCD.....	41
FIGURE 7 PERCENTAGE OF REGISTERED PATIENTS AGED 18+ WITH A DIABETES CODE (IN PAST OR PRESENT YEAR) AND REGISTERED AND ALIVE IN EACH YEAR 2000-2010, BY PATIENT'S CODE CATEGORY IN GPRD.....	41
FIGURE 8 THE PROPORTION OF DIABETIC RETINOPATHY IN THE KNOW DIABETIC POPULATION IN 2009, 2010 AND 2011 (NPCD).....	42
FIGURE 9 HBA1C METRICS BETWEEN YEARS AND WITHIN PRACTICES IS CALCULATED FOR 2009, 2010 AND 2011 (NPCD). VERTICAL BARS SHOW UPPER AND LOWER LIMITS OF PRACTICE COMPLETENESS, CORRECTNESS AND ACCURACY, 25% AND 75% QUANTILES AND MEDIANS.....	43
FIGURE 10 THE PROPORTION OF PATIENTS WITH TYPE-2 DIABETES WITHIN A GIVEN PRACTICE HAVING AT LEAST ONE DIAGNOSTIC LAB RESULT (HBA1C) IN 2009, 2010 AND 2011 IS COMPUTED. PRACTICE IS HERE THE QUALITY DETERMINANT.....	43
FIGURE 11 BMI METRIC BETWEEN YEARS AND WITHIN PRACTICES IS CALCULATED FOR 2009, 2010 AND 2011 IN NPCD. VERTICAL BARS SHOW UPPER AND LOWER LIMITS OF PRACTICE COMPLETENESS, CORRECTNESS AND ACCURACY, 25% AND 75% QUANTILES AND MEDIANS.....	44
FIGURE 12 THE PROPORTION OF PATIENTS WITH TYPE-2 DIABETES WITHIN A GIVEN PRACTICE HAVING AT LEAST ONE BMI MEASUREMENT IN 2009, 2010 AND 2011 IS COMPUTED.....	45
FIGURE 13 THE PROPORTION OF PATIENTS WITH TYPE-2 DIABETES WITHIN A GIVEN PRACTICE HAVING AT LEAST ONE MEASUREMENT OF A GIVEN TYPE WITHIN A GIVEN YEAR.....	46
FIGURE 14 CONCEPTUAL ARCHITECTURE.....	48
FIGURE 15 QUALITY METRICS TABLE.....	49
FIGURE 16 WORKFLOW FOR SELECTING PRACTICES.....	50
FIGURE 17 AUTHORISATION PAGE.....	51
FIGURE 18 BACKGROUND INFORMATION ABOUT THE AVAILABLE DATABASES.....	51
FIGURE 19 SELECTION OF GENERAL REQUIREMENTS FOR THE DATABASES.....	52
FIGURE 20 SELECTION OF DATABASE.....	52

FIGURE 21 GENERAL INFORMATION ABOUT THE SELECTED DATABASE	52
FIGURE 22 SELECTION OF THE USE CASE	53
FIGURE 23 SELECTION OF THE YEAR	53
FIGURE 24 DATA ELEMENTS FOR THE USE CASE.....	53
FIGURE 25 DISTRIBUTION OF HBA1C METRICS OVER PRACTICES IN NPCD IN 2010.....	54
FIGURE 26 RESULT OF THE QUERY	54
FIGURE 27 POSSIBILITY TO SAVE THE QUERY OR SEND A REQUEST TO THE DATABASE	54
FIGURE 28 FEEDBACK ABOUT THE DATA QUALITY METRICS	55
FIGURE 29 PERFORMANCE EXPECTANCY EVALUTION	57
FIGURE 30 EFFORT EXPECTANCY EVALUTION.....	58
FIGURE 31 THE EFFECT OF SOCIAL INFLUENCE.....	59
FIGURE 32 FACILITATING FACTORS	60
FIGURE 32 BEHAVIOURAL INTENTION CONSTRUCT FOR THE DQT TAM EVALUATION.....	62
FIGURE 33 PERCENTAGE OF NON-TEMPORARY PATIENTS AGED 18+ WITH A CODE FOR GORD IN 2010 IN NPCD AND GPRD.....	70
FIGURE 34 PERCENTAGE OF GORD PATIENTS RECEIVING NSAIDS AND PERCENTAGE OF PATIENTS WITH HEARTBURN OR ACID REGURGITATION RECEIVING PPI IN 2010.....	70
FIGURE 35 THE PROPORTION OF PATIENTS WITH GORD WITHIN A GIVEN PRACTICE HAVING AT LEAST ONE MEASUREMENT OF A GIVEN TYPE IN 2010 CALCULATED FOR NPCD AND GPRD.....	71
FIGURE 36 ESTIMATES OF COMPLETENESS FOR A VARIETY OF DATA ENTITIES FOR THE POPULATION OF PATIENTS WITH GORD IN NPCD OVER THE YEARS.	71

List of Tables

TABLE 1 <i>DEFINITIONS FOR COMPLETENESS, ACCURACY AND CORRECTNESS AND CONSISTENCY.</i>	23
TABLE 2 <i>CHARACTER AS DEFINED IN THIS REPORT AND POSSIBLE EQUIVALENT TERMS IN THE LITERATURE.....</i>	24
TABLE 3 <i>THE TWO BASIC OPERATIONS PROPOSED FOR THE DATA QUALITY MODEL: COUNTING AND ENUMERATION.</i>	26
TABLE 4 ALL DIAGNOSIS CODES FOR THE DIABETIC POPULATION OF EACH PRACTICE ARE ENUMERATED AND COUNTED AND THEIR POSITION WITHIN THE ICPC CLASSIFICATION DETERMINED. HERE WE ONLY SHOW 3 CODES FOR DIABETES. THIS ALLOWS FOR AN ASSESSMENT OF CORRECTNESS AND ACCURACY (AGAINST SCIENTIFIC CONTEXT). HERE, THE CODING IS ACCURATE AND THE PROPORTION OF TYPE-1/TYPE-2 DIABETICS CONFORMS TO SCIENTIFIC UNDERSTANDING AND APPEARS CORRECT.	26
TABLE 5 CATEGORIES OF FACTORS THAT INFLUENCE THE GENERATION AND TRANSMISSION OF DATA IN DIFFERENT ZONES OF THE TRANSFORM ZONE MODEL (35).....	31
TABLE 6 THE APPLICATION OF THRESHOLDS TO EACH METRIC PROVIDES A SPECIFICATION OF DATA QUALITY FOR A STUDY.	32
TABLE 7 BASIC CHARACTERISTICS OF NIVEL PRIMARY CARE DATABASE AND GENERAL PRACTICE RESEARCH DATABASE.....	36
TABLE 8 DATA ELEMENTS OF INTEREST FOR THE DIABETES USE CASE	39
TABLE 9 QUALITY METRICS FOR DIABETES USE CASE.....	40
TABLE 10 SUMMARY RESULTS FOR THE FACILITATING FACTORS CONSTRUCT OF TAM.	60
TABLE 11 SUMMARY RESULTS FOR THE EVALUATION OF THE SPECIFIC FEATURES OF THE DATA QUALITY WEB TOOL.....	61

TABLE 12 SUMMARY RESULTS FOR THE USER'S BEHAVIOURAL INTENTION IN THE FUTURE
ADOPTION AND USE OF THE DATA QUALITY WEB TOOL..... 62

TABLE 13 DATA ELEMENTS OF INTEREST FOR THE GORD USE CASE..... 68

TABLE 14 QUALITY METRICS FOR GORD USE CASE 68

Executive summary

This deliverable entails:

1. a generic method to assess primary care EHR data quality by quantifiable means.
2. A software tool to assess data quality in a given dataset.

In section 2 we give an overview of the different purposes that EHR data may serve and the factors that influence the extent to which this data can be used for these purposes. The main issue discussed here is that the different purposes for which data are collected have different effects on the quality of the data. The very idea of data being collected to support enhances the need to assess the quality of any given dataset before it gets used, bearing in mind the background the different steps that need to be taken from an event to take place to the calculation of outcomes based on the occurrence of this event in a database.

In section 3 a general data quality framework is presented that acknowledges the fact that quality metrics require a given purpose and that any data quality metrics should be described solely in terms of accuracy, correctness, completeness and consistency. In this section we present an axiomatic approach to data quality metrics and dimensions, leading any research project through the appropriate choices regarding data quality metrics.

In section 4 we take the TRANSFoRm diabetes use case as an example to implement the framework described in section 3.

Section 5 gives a brief description of the data quality tool that was prototyped during the project, which allows researchers to visualise the quality of the data in a given database and to make practice data selections based on requirements set by a researcher. To our knowledge this is the first attempt to give researchers full control of relevant quality metrics to select practices for a given research project. In the actual data quality tool researchers can vary their desired cut off values data quality standards on a selection of quality indicators by graphic sliders and have an idea about the number of subjects and practices that would result using these cut-off values. Evaluation took place using a questionnaire based on the Technology Acceptance Model. The over-all responses were positive, and the Data Quality tool seems likely to enhance the use of routine EHR data for research purposes.

1 Introduction and structure of the report

In many European countries general practitioners are responsible for first contact, continuing, and generalist care of the entire population from birth to death. Even in countries where general practitioners do not fulfil a 'gatekeeper' function, controlling access to specialist services, much of the future course for an individual patient with a health problem is determined in primary care. As computerisation of primary care facilities is rapidly increasing, a wealth of information is created in routinely recorded electronic health records (EHR) that can be and is used for a wide range of purposes.

The central theme of this report is data quality, which we describe as whether the data is 'fit' for a given purpose.

In section 2 we give an overview of the different purposes that EHR data may serve and the factors that influence the extent to which this data can be used for these purposes.

In section 3 a general data quality framework will be presented that acknowledges the fact that quality metrics require a given purpose and that any data quality metrics should be described solely in terms of accuracy, correctness, completeness and consistency.

In section 4 we focus on the diabetes use case and its quality metrics.

Section 5 gives a brief description of the data quality tool that was prototyped during the project, which allows researchers to visualise the quality of the data in a given database and to make practice data selections based on requirements set by a researcher.

2 Purposes of EHR data and factors influencing data quality

According to van der Lei et al. data should only be used for the purposes for which they were collected (1). Any academic endeavour on data quality should therefore start with a critical analysis of the purposes with which data were originally collected and how this purpose may influence the quality of the data for other purposes. In this section therefore, we give an overview of different types of primary and secondary use and how different types of use affect the quality of the data being recorded.

2.1 Purposes of using primary care EHR data

Primary use: patient care

First of all, electronic health records are used to manage an individual patient's care. Many patients receive health care from a wide variety of health care providers, and for the involved care providers to share relevant information on patient's health problems and treatment is becoming increasingly important. To share this information in a useful way, this information needs to be complete and correct, relative to care needs. The more people use the information, the more serious the consequences of incomplete or incorrect information on for example (co) morbidity, intolerances, co-medication and allergies.

Also the use of decision support systems is rapidly increasing. More than 20% of Dutch general practices use the decision support system expert.doc(2). For almost all Dutch GP practices use the evidence based electronic prescribing system (electronisch voorschrijfsysteem) for prescribing drugs. To use these support systems in a proper way, the underlying data need to be complete and correct.

Secondary use: Management information, quality indicators

EHR-data is increasingly used for quality of care indicators that are used for management information within the health care facility itself, but also as a source of information for third party organisations such health insurers or governmental bodies. This is not always unproblematic (3). For example, Stirbu-Wagner and coworkers found in the Netherlands in 2008 that it was hardly possible to retrieve the necessary data from EHR systems (4). Technically the necessary data elements could be extracted from the EHR systems, but the quality of the data was poor. However, the situation regarding data quality in the Netherlands is likely to have changed in recent

years. Substantial numbers of practices nowadays receive feedback on the quality of recording, based on the EHR-scan (5) and in 2012 and 2013, part of the reimbursement scheme of GPs is based on the quality of recording. In the UK's Quality and Outcomes framework (QoF), GPs earn a large part of their income based on the data about the quality of the delivered care based on EHR data. Experience in the United Kingdom showed that quality of recording improves considerably if the finances are made dependent on it. This is true, at least, for the items that generate reimbursement (6).

Secondary use: Research

Increasingly, repositories based on EHR data are used for research purposes, including epidemiological studies, health services research and clinical trials. The use of EHR data for scientific research has several important advantages compared with surveys as EHR data suffer less from systematic errors such as selective non-response, response bias¹ and recall bias². Also, EHR-data are collected routinely, which makes it less expensive than surveys. A further advantage is that EHR data are being recorded continuously, which allows researchers to construct longitudinal data files to evaluate for example the effects of an event that could not have been foreseen. This applies not only to environmental hazards (7) but also for reviewing health care reforms (8);(9). EHR data provide information about how the health care system is functioning and about developments in public health at comparatively low cost. Linking this data to other data sources increases the research possibilities enormously. Data from NIVEL's primary care database have been linked to a large number of other sources such as living environment (10), ethnicity (11), income, school dropout, insurance claims and pharmacies (12). EHR data are also frequently used for public health forecasts and surveillance (13). A rather new but upcoming type of research using routine data is clinical trials, yet another type of research that is carried out using electronic health records (14;15).

However versatile EHR data can be, for each of the purposes mentioned to deliver useful results, an assessment needs to be made to what extent the data are actually fit for each of the above purposes.

2.2 Factors influencing data quality: a stepwise approach

There are a number of reasons why data quality may not be up to the required standards involving all the steps that lead from a clinical event, to the end

¹ Systematic error caused by social desirability or leading questions.

² Systematic error caused by differences in the accuracy or completeness of the recollections of events or experiences from the past

user who finds this event in a research file and starts using it in a query. This whole process is depicted in Figure 1.

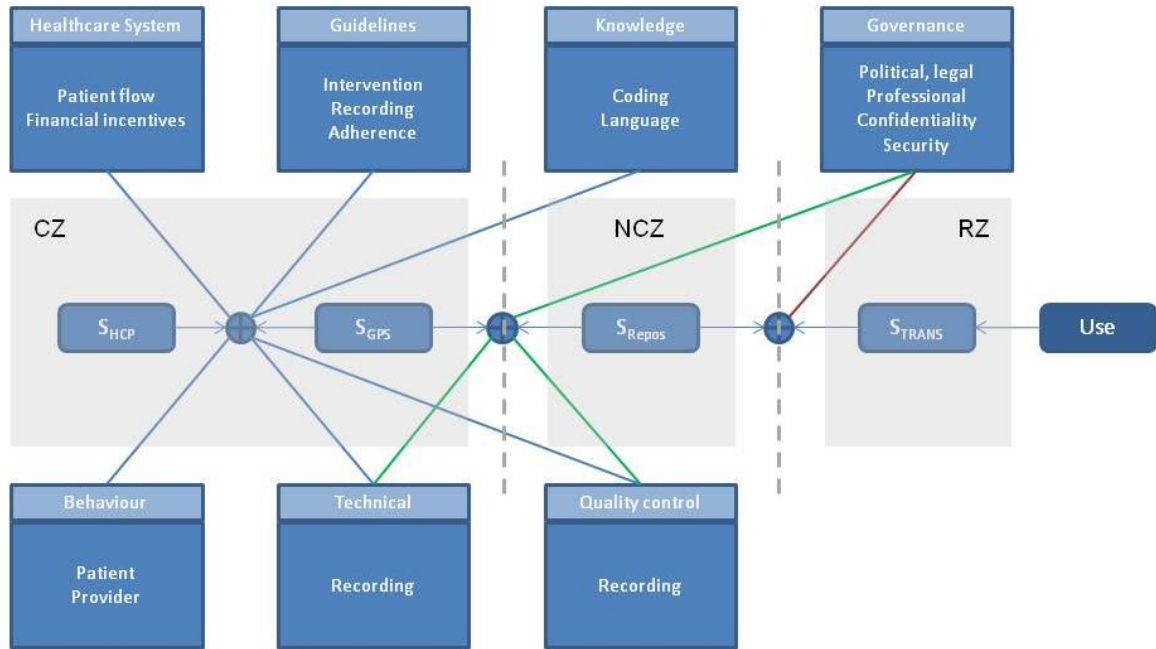


Figure 1 The overall process from patient access and measurement in a clinical setting S(HCP) to end use by the researcher (Use). CZ – care zone; NCZ – non-care zone, or linkage zone; RZ – research zone. S(HCP) – Health Care Professional; S(GPIS) – GP Information System; S(Repos) – Clinical data repository system; S(TRANS) – TRANSFoRm system; Use – End User.

Step 1: event must take place. This may seem trivial, but for a BP reading to be recorded in an EHR system and finally in an EHR based repository, a patient’s BP needs to be measured first. The likelihood of such a measurement to take place is dependent on a number of factors:

- There can be different types of incentives to stimulate GPs to perform certain tasks.
 - Organisational aspects of the health care system. Here the difference between gatekeeping systems and non-gatekeeping systems is important. In a non-gatekeeping system, a BP reading may take place outside primary care, resulting in fewer BP readings in primary care settings. Similarly, the existence of a list system, where people are listed with a particular practice may not affect the number of BP readings in primary care as a whole, but it will affect the number of BP readings with a particular doctor.
 - Financial incentives in the health care system. The reimbursement system in one country may stimulate BP

readings, whereas in other countries it will not. In the Netherlands, BP readings are supposed to take place every year in the diabetes population, whereas different rules apply for other segments of the practice populations. The Quality and Outcomes framework in the UK promotes the whole population to get a BP reading every year (6). Incentivisation in the health care system may affect completeness of the data.

- Quality of care guidelines describe what a GP normally should do given a certain condition. If a guideline says a blood pressure reading should be done every year, it will be more likely that such a measurement takes place. Quality of care guidelines may affect completeness, correctness and accuracy of data.
- Practice workload may have a negative effect on the collection of electronic data where the number of staff in the practice and the consultation length don't offer the opportunity. Practice workload may affect completeness, correctness and accuracy of data.

Step 2: event must be recorded. Each event must be recorded in order to eventually be available in a repository. Whether an event gets recorded is again dependent on a number of factors:

- Computerisation and EHR software. Although all software packages in the Netherlands and probably also in other countries are qualified as 'good' software packages by the National ICT institute, a recent study in the Netherlands showed that there are considerable differences between software packages regarding the user friendliness with respect to recording items like contra indications and the way episodes of care are recorded (5). Functionalities of the EHR software may affect completeness and correctness of the data.
- Strategic recording behaviour. Two different mechanisms may be at work here.
 - Monetary incentives inherent to the health care system may lead to strategic recording behaviour. Separate reimbursement schemes for patients with chronic illness will stimulate GPs to diagnose patients with chronic disease. Upcoding has been found to be a risk in relation to diagnosis related groups as a basis for reimbursement (16). Strategic behaviour may lead to incomplete, incorrect and inaccurate data.
 - Sharing of information may also have an effect on whether an event gets recorded. For example, GPs may be more reluctant to record a diagnosis of which they are not completely certain in situations where this information is shared with colleagues. This

effect may vary dependent on the event that is involved. A GP may for example be more hesitant to record depression as a diagnosis than diabetes. Similarly, GPs will be more hesitant to record for example excessive alcohol intake if this information is shared with other professionals. Yet another effect may arise from the fact that in some situations also patients will be able to see what has been recorded by the GP, but with similar results in the sense that a GP will be inclined not to record specific things that might be relevant for clinical practice, but that can cause problems in the relation with patients. Sharing information may lead to incomplete and/or incorrect data.

- Recording guidelines. In some health care systems there may be clear guidelines as to what should be recorded by GPs in an EHR (17); (18); (19). In other countries such guidelines may not exist. Absence of recording guidelines will lead to less accurate, less complete and less correct data.
- The use of coding systems and free text. Coding systems will affect what can be recorded. For example the variety of ICPC codes is much smaller than Read codes or Snomed. This will affect what can be recorded. Furthermore, the use of free text is generally regarded problematic and is only useful for small scale studies, unless this free text can be turned into data that can be processed automatically (20). Within the European context this difficulty is magnified by the presence of many languages and coding systems. Where free text is used the difficulty of establishing a clear concept or concepts for the entry will be increased. Some coding systems allow for more accurate recording than others.
- Knowledge. Software packages and coding systems may enable GPs to do all that is required and guidelines may tell him what to do, but if a GP does not know about them, this is all useless. Sub-optimal use of the EHR system will lead to incomplete or incorrect data.
- Practice workload. Shortage of time in a consultation will not enhance proper recording behaviour.

Step 3. Data must be extracted. Unless data are only used within the recording practice (the care zone), it needs to be extracted and transported (or otherwise made viewable) at another site. Factors that affect the quality of the data here are:

- Extraction software. It is the extraction software that determines what data elements are extracted. Different extraction tools, working in combination with different EHR systems will render different results in terms of data quality. This may lead to incomplete and/or incorrect data.

- Governance issues. Some patients will object to using ‘their’ data for sharing with other health care professionals or for research. Similarly, some practices will allow the use of ‘their’ data and others will not. Governance may lead to incomplete data for some patients and/or the absence of data from some practices.

Step 4. Extracted data must be imported into a database. It may seem trivial to say that all data that is extracted also needs to be entered into a database, but it is not. Whether extracted data is actually imported into a database is dependent on:

- Capacity of a database to capture the data that is extracted. This is particularly important in cases where data come in multiple formats into one single database. Semantic interoperability is the key word here (21). Data formats and even the meaning of otherwise very similar data elements may be different across software packages, and may also vary over time, dependent on for example changes in the reimbursement system.
- Quality control measures can be employed before data is entered into the database. This may for example mean that data that are incomplete are not entered into the database.

Step 5. Generating a research data file. Normally, researchers do not do their analyses on raw data or repository data, but on a dataset that is derived thereof. Not all variables in a database may be relevant for a particular study and may be excluded from the research datafile.

- Selection of data. Quality checks or filters can be employed before or after data is entered into the database. This means that not all data that is in a repository will go into a datafile that is used by a researcher.
- Linkage studies. Where data is linked, the resulting database will hold only data on the common population. This will affect completeness of the data. Complete data will only be available from the population that the two (or more) linked datasets have in common.
- Governance. A repository may not be able to facilitate all types of research. There may be regulations and steering committees that will or will not grant the possibility to use a certain repository for a certain purpose. This will affect the completeness of data.

Step 6. Analysis of the data. Different researchers will make different choices with respect to the analysis of the data. Different methods render different results. Reeves and coworkers found that different methods for computing quality of care scores can lead to different conclusions (9). This

may seem like a truism, and perhaps it is, but the fact that data isn't the same as information and that there is analysis in between that may lead to incomparable outcomes, is often overlooked.

2.3 Conclusion

We started this section with Van der Lei's statement that data should only be used for the purposes it was collected and showed that there is a variety of purposes for which EHR data are collected and that these purposes can have important effects on the quality of the data. This implies that any user of this data should be aware of the origins of the data and how this may affect the quality of the data.

3 Data quality framework

3.1 Introduction

For primary uses of data, the importance of well-designed data collection procedures is understood. These design considerations range from careful selection of study subjects to control for bias; establishing conditions for physical measurements to ensure reproducibility and consistent interpretation; the selection of questions for inclusion in questionnaires; whether such measurements are performed by study subjects or research personnel; organisational measures to ensure completeness of recording; how missing data is handled; any transformations to the raw data before storage; the recording of data provenance in general; unambiguous identification of

collected data; who may collect and correct data; and training and selection of study personnel for all of the former. Such procedures are normally fully documented in a set of Standard Operating Procedures (SOPs) for a study, and should ensure high quality data. In other words, the issue of data quality and data provenance is recognised and dealt with at the beginning of the study and adequate measures can be taken to limit problems with imperfect data.

For research based on large quantities of data that have already been collected for a different purpose, these design criteria should generally not be relaxed, or be relaxed to only a small extent. The onus still rests on the investigator to establish that the routine data forming the study have been collected to the required degree of quality. This is particularly true when the sources of data are expected to be heterogeneous, as in primary care practices throughout Europe.

Such a data quality assessment can take place based on our knowledge of the subjects' condition (population definition), the provenance of the data, and its consistency through computational checks on the data. For example, suppose we require complete blood pressure (BP) data from included practices over a 6-month period. Inconsistencies over time and practice might indicate a problem in completeness. Additional administrative data might take us further. For example, if the specific measurement device is known and the member of staff supervising the measurement is known further consistency checks can be performed, which might suggest calibration or procedural errors. These errors cannot automatically be assumed, but must be judged, since the particular subjects processed by those specific devices and staff members may look different from each other. However, sufficiently large numbers of measurements and subjects should rule out this possibility.

Generally however, extensive supporting administrative data is not available to establish context, and consequently only an understanding of the factors affecting data collection can give us confidence in the data being used. Therefore, where BP data from across Europe is being integrated, the provenance of such data must be investigated before use. Computational checks are always a starting point for this investigation and can be annotated with provenance information.

This difficulty in fully establishing context makes data quality assessment a very difficult problem and resource intensive. Many of the sources of contextual ambiguity are likely to be systematic (as in staff procedures and guidelines) and consequently large amounts of data do not necessarily

circumvent the problem. Efforts to detect such systematic errors are therefore vital.

In initial attempts at re-use of routine data an understanding developed of particular sources by the related research and data analyst community. Methods of assessing data quality, while valid, may not have been formalised, and would often have been incomplete being directed at particular areas of interest. The moves to access heterogeneous data in a wider range of contexts – geographically and institutionally - by a more disconnected research community has introduced the need for a more formal understanding and approach to assessing data quality in data that have been collected for a different purpose (22).

Two statements are frequently cited regarding data quality: “*data should be used only for the purpose for which they were collected*” (1) and “*data are of high quality if they are fit for their intended uses in operations, decision making and planning*”(23).

The first of these (by Van der Lei) must be rejected if we are to entertain the very idea of re-use of routine data and contradicts the basic idea of data re-use. The second (by Juran) does offer a starting point for a framework. It does acknowledge the possibility of data re-use, but it acknowledges the fact that this data may not always be fit for purpose and that data quality must not be taken for granted. Juran’s statement can be seen as an anachronistic warning to all those who believe that sufficiently large and diverse amounts of data will allow us to disregard the quality and provenance of that data. More data does not substitute for better data.

However, Juran’s statement does not offer any guidance on how we might operationalise an assessment of data quality, and in particular how we might make it amenable to calculation. Literature reviews have shown that attempts to systematise data quality assessment have led to a wide variation in nomenclature and conceptual framework and the roles played by truth and judgement.

We searched for existing literature on data quality, focusing on reviews published in peer reviewed journals as well as examples of existing data quality frameworks used in practice. Even though there are good examples of data quality frameworks in operation in the UK (17) and Canada (24), and recording guidelines are implemented in the Netherlands (25) and Belgium (26), concepts of various kinds remain mixed together in the literature and there has been no attempt to take an axiomatic approach (3;27-30). We seek to rationalise and clarify what work has been done before and cast this in a simpler proposal that is amenable to calculation. The latter aspect is crucial,

as the purpose of WT5.1 was to establish a procedure for assessing the quality of data available to the TRANSFoRm platform. This procedure includes both human judgement and computable operations that together permit the assessment of the quality of a given dataset using the data quality webtool.

3.2 The role of scientific or managerial context

We assert that any assessment of the quality of some particular data must take place within a scientific or managerial context. For example, the records of a patient with type-2 diabetes should correspond to our understanding of that disease and care process from the perspective of aetiology, presenting symptoms, diagnosis, and prognosis through to treatment and outcome. If this were not the case, then we must likely conclude that there are quality failings in the overall process, whether from medical or informatics professionals, professional groups, or the systems themselves. Indeed, turning this in reverse, for data that is taken to be of high quality, anomalies in the data may point to new knowledge. Sometimes this may be problematic as part of the background knowledge may be derived from the data under study (eg. GPRD and NIVEL data are used to generate public health information).

For example, for the population of patients with diabetes in Europe and North America it is known that the ratio of type-1 to type-2 diabetes is about 10%; their BP is raised above that of the general population; they are very likely to have had an HbA1c measurement made whereas the general population will not; and their outcome to treatment may depend on specific hereditary factors. The care process itself will be reflected in understanding of a similar kind.

While this specific understanding is crucial to quality assessment the more general scientific (and managerial) context is also useful, and can be found encapsulated in international and national classification schemes, such as ICD10 or SNOMED-CT, or in population surveys of things such as height and weight, or in more general care processes. These types of understanding also play a role.

It is important to note that the truth of a data value or values can never be known, even where comparisons are available, as comparators are likely drawn from the same initial measurement process. Therefore, appeals to truth should not be part of any formalism. What can be assessed are the values' conformance with the scientific or managerial context and this can be tested

in many ways, making clear the role of judgement alongside computable aspects of quality assessment. For example, the medication metformin is almost exclusively given to those patients diagnosed with type-2 diabetes and only after an HbA1c measurement has been made. This temporal correlation between these two data variables can be tested. There are no limits to these kinds of assessments of the data.

What this makes clear is that data quality can only be discerned as a discrepancy between the data (and the information it provides), and the scientific and managerial context within which it was created and transmitted.

3.3 Axiomatic basis – purpose and population

The literature on the subject of data quality clearly shows a difficulty in grounding the conceptual discussion of data quality, with a proliferation of concepts and associated terminology (28). It is the intention here to produce a framework with only a few powerful base concepts and well-defined terminology, with the aim of establishing computable metrics for quality.

The most important concept associated with that of *data quality* is *purpose*. Both of the statements cited above (1) (31) link *quality* and *purpose* in a single loose definition or rule. The former statement implies that good data quality must be engineered prior to data collection, while the latter suggests that good data quality can be obtained after data collection through judicious use of the data. If routine data is to be re-used (with the economy that brings to data acquisition) then we must reject van der Lei and embrace Juran. It should also be clear that purposes can vary greatly in their specificity, but we assert that there is nothing fundamental to distinguish general from specific purposes in any way. Therefore, we offer the following as the first axiom of our framework:

A1: *Data quality* can only be determined within the context of a specified *purpose* and related *science and management*, and that purpose must be explicitly stated (be well-specified)

This purpose is realised as a set of variables of interest and the use to which they will be put. In considering where such data might be found, the concept of population is fundamental. We would therefore offer a second axiom:

A2: A well-specified *purpose* must determine a *population* of interest

These concepts and axioms are shown in Figure 2, which makes clear the primacy of context, purpose and population (availability of data).

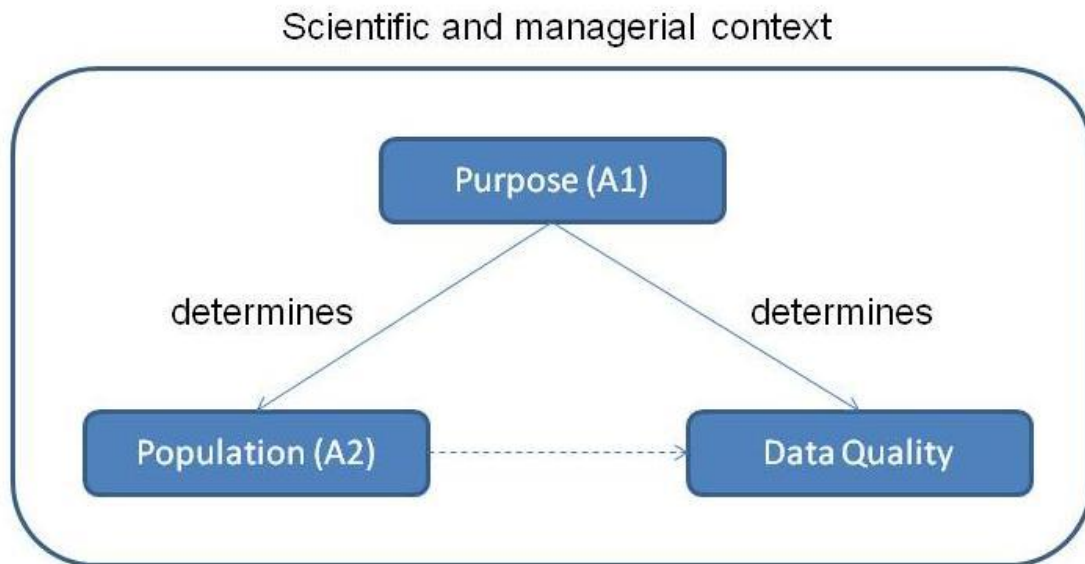


Figure 2 Illustration of the relationship between the core concepts of the data quality model. A1 – axiom 1, A2 – axiom 2.

3.4 Population determination

The specification of a population is necessary to define ‘where’ the data of interest can be found, and the size of this population will define the denominator for the quality metrics we shall describe later. We must take care to make sure that all the chosen sources of data determine comparable populations and that this determination takes into account the scientific and managerial context discussed in the previous sections. Each data source must decide the best criteria for establishing this comparable population and as there are variations between sources in how, where and why data is collected this will vary from source to source. Such population algorithms are an important area of research (32) and their specification involves a similar quality assessment to that of the study data. Such an assessment may make use of similar metrics to that for the study data, but may not require similar strength results to establish a comparable population. For example, a study may require complete data on HbA1c and drug consumption to be of sufficient quality. On the other hand only a few laboratory measurement, drug records or diagnosis codes will suffice to establish a patient with type-2 diabetes with

reasonable certainty. There may be purposes that define the whole population listed in a practice. An example of this might be the purpose of generating incidences and prevalences for the whole population. In that case the quality of diagnosis recording in the whole population is relevant. This may result in what might be referred to as ‘general quality indicators’ such as discerned by Tate et al. (33).

3.5 Measurement Value Set (MVS)

In the model presented in Figure 2, the data quality concept has been left unexplained. However, it is reasonable to suppose that any statement of quality must be made against one or more data variables, such as HbA1c, that are available from a population within a particular data source. The data variable in this case is realised as a finite set of values drawn from all possible HbA1c measurements. We designate this set of values in the selected population and source as the *measurement value set* for HbA1c.

When dealing with many heterogeneous data sources we must specify the equivalent variables in these sources in a more formal way. Variables can be expressed unambiguously in TRANSFoRm through the dual use of the Clinical Data Integration Model (CDIM, WT6.5) and a chosen reference terminology such as LOINC or SNOMED-CT, and their expression may make use of multiple CDIM concepts and coding systems. In general, these formal definitions of variables are dependent on one or more ontologies.

For example, the variable ‘HbA1c by IFCC method’ may be specified as:

Req/Resp	CDIM concept	Example value(s)
Request	Laboratory_test_ordering	LOINC: 59261-8
Request	Measurement unit label	UCUM: mmol/mol
Response	Scalar measurement datum (laboratory finding)	53

where *laboratory_test_ordering* is a concept from the CDIM, LOINC is a coding system for laboratory measurements, and UCUM is an international standard for units of measurement. These three attributes unambiguously identify the variable and allow sources to determine whether their measurement value sets correspond to this definition.

Similarly, the variable ‘diagnosis of diabetes mellitus type-2’ may be specified as:

Req/Resp	CDIM concept	Example value(s)
Response	Diagnostic_conclusion	ICD10: E11

Scientific studies have inclusion and exclusion criteria as well as outcomes and covariates. Measurement value sets would be obtained for all these variables. For example, in the context of a study on diabetes we would obtain value sets for each of: age, gender, and ethnicity, date of diagnosis of diabetes, other diagnoses, various laboratory measurements, medication, vital signs, lifestyle and family history.

3.6 Characterisation of a measurement value set

In the context of data quality, what can we know about a variable’s value set? In general, we know that the content of the value set relates to an overall process from patient access and measurement in a clinical setting through to recording on an EHR system and its subsequent transmission through further systems (Figure 1). In any given setting this process may be disturbed in a variety of ways. For example, a blood sample from a patient may take too long to get to the laboratory for testing, resulting in altered serum potassium levels; a clinician may measure a blood pressure without reference to guidelines; a clinician may measure a blood pressure but fail to record this within the EHR; a clinician may record a diagnosis to a reduced degree of accuracy, e.g. diabetes mellitus rather than diabetes mellitus type-2. However, these disturbances are not known directly and so any assessment of the value set only has the values themselves to work with.

As a third axiom we shall now state:

A3: The measurement value set is fully characterised by specifying its *completeness*, *accuracy* and *correctness*. No further characterisation is necessary.

We define the *character* of data as the degree to which it is *complete*, *accurate* and *correct* and we provide definitions below for these terms (Table 1). In the literature, character is more commonly termed *dimension* (34). All dimensions discussed in the literature can be shown to be essentially equivalent (see 2 where we map these three characters to the dimensions found in the literature) or relate to the operations described in section 3.7.

Character	Definition
Completeness	On the completion of a measurement or judgemental process, the <i>degree</i> to which the outcome of the measurement or judgement is recorded in the EHR and subsequent systems along with the necessary spatial and temporal context. (The spatial context is logical rather than geographical). For example, if a physician completes a measurement of systolic blood pressure within general practice, it is the probability that that measurement will be recorded in the EHR along with the clinical setting (space) and time of measurement. Transmission to further systems may reduce the level of completeness.
Accuracy	The <i>degree</i> of detail to which a measurement or judgemental process has been taken. For example, a patient's height may be recorded as 1.6m, but more accurately as 1.63m. A diagnosis may be recorded as diabetes mellitus (ICPC: T90), but more accurately as diabetes mellitus type-2 (T90.02).
Correctness	The <i>degree</i> to which the recorded value conforms to the scientific and managerial context. For example, does a diagnosis code conform to the context? Do other codes provide supportive evidence?
Consistency	The <i>degree</i> to which each of the characters of correctness, accuracy and correctness are unchanged over time and space, where time refers to the moment at which these characters are assessed.

Table 1 Definitions for completeness, accuracy and correctness and consistency.

These characters and their definitions are distinct from the operations that might be used for their calculation – hence the use of the word *degree* in the definitions.

It is instructive to map some of the *dimensions* listed in the literature (34) to the *characters* defined above. All dimensions appear able to be mapped. However, some dimensions from some authors appear to mix more fundamental dimensions. For example, accurate has been used by some to mean complete and correct.

Character	Dimension(s) mentioned in the literature
Completeness	Completeness, availability, missingness, omission, presence, sensitivity
Accuracy	Granularity, level of detail
Correctness	Error, validity, plausibility, believability, trustworthiness, positive predictive value, agreement, concordance.
Consistency	Reliability

Table 2 Character as defined in this report and possible equivalent terms in the literature

A number of dimensions mentioned in the literature appear to be absent: *currency*, *recency* or *timeliness*. We do believe these terms are important. We believe, however, that timeliness is covered by completeness (if data are not delivered, this will result in incomplete data) on the one hand, and more related to service quality than to data quality. For example, data does not change its quality depending on whether it arrives at a point of analysis or deliberation after 1 second or 10 days or two years. It may result in incomplete data for a given time period and is therefore covered by the term completeness. The mixing of dimensions is prominent within the literature and is a source of confusion.

3.7 Metrics

A *quality metric* is defined as a numeric evaluation of one of the characteristics - completeness, accuracy, and correctness – computed through either a *counting operation* or an *enumeration operation* on a measurement value set for a variable of interest.

We also introduce the definition of a *quality determinant* as a parameter that may influence the value of the metric. For example, practice settings will influence the quality of data collected through financial incentives, adherence to guidelines, and physician behaviour. We would therefore clearly expect our metrics to *vary by practice*. Periods of time are also quality determinants. For example, the introduction of a financial incentive or mandate will result in improved completeness of some data from the time of introduction. As in the case of the practice setting the primary determinant is the introduction of an incentive, but the measureable determinant is time.

The counting and enumeration operations we propose are listed below:

Operation	Applies to character	Example metric
<p>Count Non-null values are <u>counted</u> for a measurement value set and expressed as a ratio to a denominator</p>	Completeness	<p>HbA1c values are counted and the counts <u>expressed</u> as ratios to the population denominators and <u>profiled</u> by a quality determinant such as practice or period of time (context). See Figure 3.</p>
<p>Enumerate (distribution) Non-null values for a measurement value set are <u>counted</u> in bins and expressed as a ratio to a denominator. For numeric quantities the bin increments are dependent on the measurement value set. For coded values (including ordinal) the bins cover the chosen code domain.</p>	Completeness	<p>Value distributions are <u>inspected</u> for missing values expected within the scientific and managerial context. For example, a particular practice may only measure heights for their patients with diabetes and children with stunted growth. Profiling the distribution with respect to the population with diabetes may give a smooth profile suggesting correctly that these are complete. However, a similar profile for the practice as a whole will clearly show incompleteness.</p>
	Correctness	<p>Diagnosis codes are enumerated and <u>cross-checked</u> against scientific context such as ICD10 diabetes codes. Counts are provided for those within the required classifications and those falling outside the classifications.</p> <p>Age values might be compared against a known profile with known bounds for a particular population.</p>
	Accuracy	<p>The enumerated values or codes are <u>cross-checked</u> against scientific and managerial context for the degree of resolution available.</p> <p>For example, codes might be compared against ICPC and designated as a branch or leaf node in the hierarchy. A numeric variable would be designated by its</p>

		significant number of digits.
--	--	-------------------------------

Table 3 The two basic operations proposed for the data quality model: counting and enumeration.

For example, we may be interested in the quality of diagnoses within a group of general practices. An enumeration might yield the following partial result (Table 4 below).

	Variable				
Practice code (quality determinant)	Diagnosis code	Count (% practice population)	Context	Node designation	
14	...				
14	E10-E14	194 (1%)	ICD10	Branch (level n)	
14	E10	1837 (9%)	ICD10	Branch (n+1)	
14	E11	19041 (90%)	ICD10	Branch (n+1)	
14	...				

Table 4 All diagnosis codes for the diabetic population of each practice are enumerated and counted and their position within the ICPC classification determined. Here we only show 3 codes for diabetes. This allows for an assessment of correctness and accuracy (against scientific context). Here, the coding is accurate and the proportion of type-1/type-2 diabetics conforms to scientific understanding and appears correct.

From this enumeration and counting the degree of accuracy and correctness can be estimated. The scientific understanding includes expectations such as the ratio of the counts for E10/E11 at approximately 10%. The completeness of E10-E14 coding over the set of practices can be estimated by establishing the distribution of counts by a quality determinant such as practice code.

It is possible to extend these metrics to two or more variables to demonstrate expected correlations. For example, the composite (diagnosis, gender) could be enumerated to demonstrate that (prostate cancer, female) yields a low or zero count. Correlation can be used to estimate completeness and correctness within a single database. For example, the proportion of patients taking metformin that have had an HbA1c measurement within the previous 12 months can be computed against the quality determinants *practice* or *time period*.

To assess the quality of data for a study as a whole, many metrics related to inclusion/exclusion criteria and outcomes would be formulated and assessed simultaneously. Many authors have chosen to describe metrics as being general or specific quality indicators, but we would suggest this distinction is not useful. In our framework these distinctions do not appear. For example, a variable such as the time of an event or some other 'key' field, often has quality indicators calculated without apparent consideration of purpose. And yet, for example, the 'missingness' of such a variable may be tightly correlated with process at the time of recording, which might be demonstrated by correlation with another variable.

When we consider data quality in the context of TRANSfoRm we are interested in estimates of quality from groups of repositories (EHRs), and not single repositories. We wish to know which practices to include in epidemiological studies or clinical trials as a group. Therefore the metrics we seek for data quality must involve some kind of comparison between practices (and time periods) and not within practices. In the metric:

% of patients with diabetic retinopathy for all the known patients with diabetes in the practice population in 2009, 2010 and 2011 (denominator is known patients with diabetes)

We are not interested in any one % value, i.e. the value for a particular practice in a particular year. If we were, we would indeed be concerning ourselves with the truth and consideration of some 'gold standard' to help us. Instead, we are interested in how that % value varies by practice and time – the quality determinants - which does not require us to know the truth (which is why appeals to truth are not part of this framework). The variation we see leads us to believe something about whether we should use the group's data or not.

A truthful database would be one that was perfect in every aspect of the data collection and subsequent processing. A database at another site might also be perfect in its data collection and processing according to its environment (contextual provenance) and it too would offer up its truth. These would be gold standards at each site. But the minute we showed variation between the sites as described above because the contexts are different, the 'gold standard' status does not help: the data are not comparable (WT6.3).

3.8 Visualising the metrics

Figure 3 illustrates how we might visualise the data summaries that the metrics produce. In this example we are interested in the completeness

(character) of HbA1c (variable). HbA1c values belonging to the value sets for each practice are counted and expressed as a ratio to the practice populations. We can see that all practices have at least 38% of their patients with type-2 diabetes having at least one HbA1c value recorded in 2009 with some practices achieving 93%. We also note that these profiles vary by EHR vendor giving us additional clues as to why the incompleteness is present. In this instance, contextual provenance relating to the systems involved might be useful to obtain.

By turning these profiles on their side and representing them as box plots, we get the visualisation shown in Figure 4. This particular plot permits the completeness of multiple variables to be assessed at the same time and hence an overall appreciation of data quality for a study. Patterns can be discerned showing what data is being collected together and how overall completeness is improving with time (which is also a quality determinant in the graph). Mixing many quality determinants in a single graph can be very powerful at exposing incompleteness.

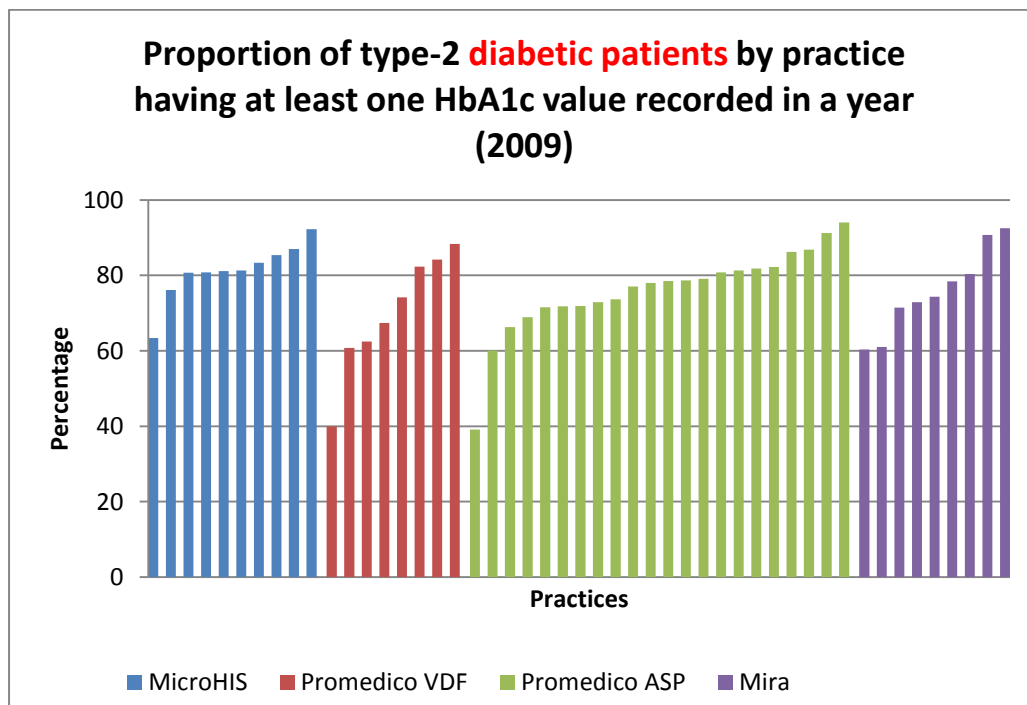


Figure 3 The proportion of type-2 diabetic patients within a given practice having at least one measurement of a given type (here HbA1c) within a given year (2009) is computed as a measure of completeness. Quality determinants in this figure include practice and EHR vendor. Taken from the NPCD database of NIVEL

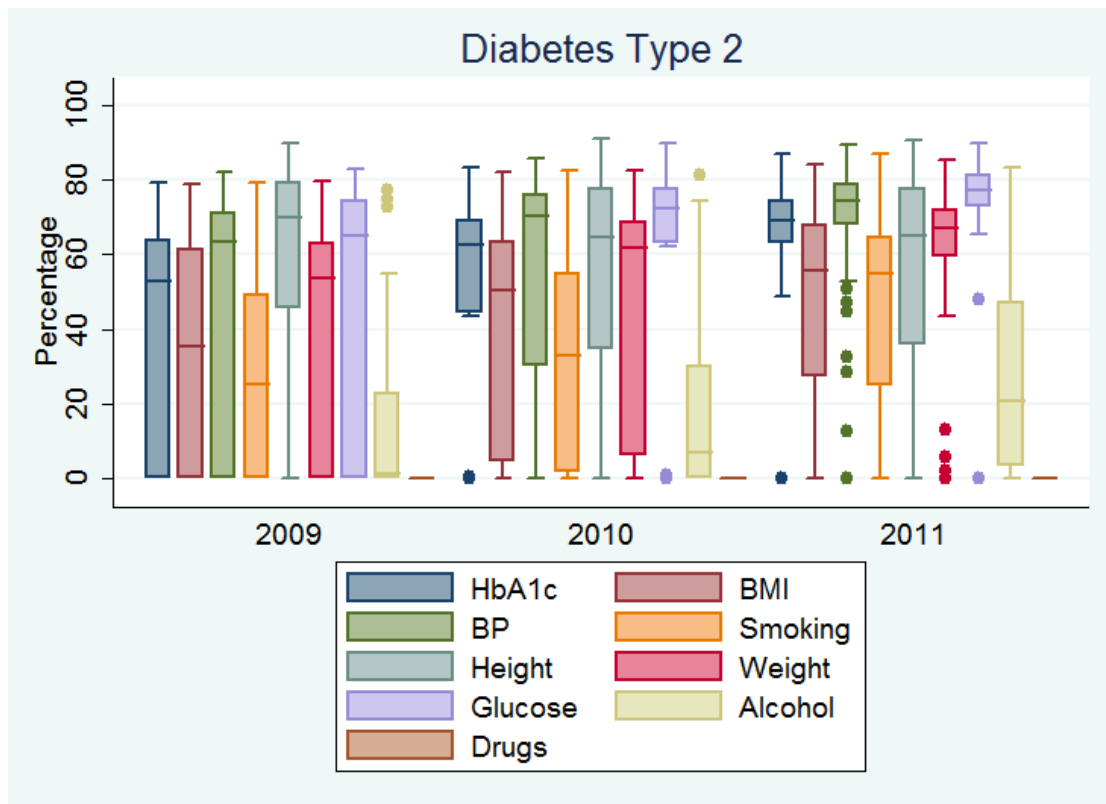


Figure 4 Box plots visualising the distribution of completeness for variables relevant for the diabetes use case.

The distribution of completeness for Individual variables (as in Figure 4) are conveniently visualised as box-plots in this example where the improving completeness of a number of data variables can readily be seen between years and within practices. The lower and upper edges of a box mark the 25th and 75th percentiles. The horizontal line within the box represents the median. The narrow boxes extending to the top and bottom are called whiskers; they extend from the quartiles to the farthest observation not farther than 1.5 times the distance between the quartiles. Any more extreme values (the outliers) are plotted with individual markers.

3.9 Contextual provenance of data

Where a variable’s metrics show a deviation from expectation or requirements, it is important to establish the reason for this before making a final judgement on whether the data are of sufficient quality. No metric result will be perfect and it is the extent of the deviation that must be considered. In making this judgement it is important to know how the data came into being: where, when and how it was generated; under what circumstances; and by what route it arrived for analysis, establishing what is termed contextual provenance (see WT6.3 and section 2). Only once the metrics and the associated contextual provenance are synthesised can a proper judgement be made on whether to use the data and state it to be of sufficient quality.

One possible approach is for metrics to be tagged with provenance information about the clinical source and intervening systems. Not only does this help underpin the metric interpretation, it offers a strategy for improvement. A full discussion of provenance and this approach was given as part of the First Project Report (D10.1). We will briefly summarise results from that report.

The provenance of data can be classified into three types – attendant data, system provenance and contextual provenance as investigated in the work on data comparability (Work Task 6.3). Attendant provenance is additional data associated with the data variable and required to make a meaningful interpretation of the data value, such as patient posture for a systolic BP measurement. System provenance is the set of parameters governing the processes which take place in the transfer of the data variables between and within systems. Both attendant data and system provenance are recorded in the form of additional data.

In contrast to these two, contextual provenance is mainly unrecorded. Examples include financial incentives for data recording or clinical guidelines. In local research, contextual provenance factors are usually known by researchers through discussion with the database owners. However, it is difficult for a researcher to obtain and appreciate these factors within international settings and a multiplicity of study requirements. In Work Task 6.3 these factors were explored and grouped into a small number of abstract categories (shown in Table 5) and related to the transaction points in the provenance model from the report (shown in Figure 1, section 2).

Factor category	Care zone (data	Non-care (Linkage) zone (ETL: EHRs to	Research zone (Data counts from
-----------------	--------------------	---------------------------------------------	------------------------------------

	recording)	repository)	platform)
Healthcare system	X		
Professional Guidelines	X		
Knowledge management	X		
Technical processes	X	X	
Quality control	X	X	
Governance		X	X
Patient/provider behavior	X		

Table 5 Categories of factors that influence the generation and transmission of data in different zones of the TRANSFoRm zone model (35).

For example, completeness of BP data may differ significantly between UK practices and Dutch practices. In the UK, practices are incentivised to collect BP data as part of their QoF returns (6) whereas the Dutch healthcare system offers this inducement only for DM2 patients and BP data is collected contingent on medical need. The accuracy of BP measurements may also vary between countries if guidelines are in place specifying this. Certain data, such as HIV status, may be entirely missing in repositories for countries where governance may prevent its release from local EHRs. By tagging metrics with contextual provenance information, barriers to research from incompleteness of data can be identified and measures taken to improve the metrics. In the report it was proposed that the researchers can assist sources with the tagging process as they request and review the metrics that sources provide. Online tools must be provided to aid this process as suggested in the report for Work Task 6.3.

3.10 Data quality statements

The judgements that arise from the metrics and contextual provenance information must ultimately be expressed as a selection over the data sources and time periods of interest - the quality determinants - and so the final step must be a statement that yields a selection against one or more quality

determinants. For example, the researcher may choose a particular group of practices, periods of time and EHR vendors that provide the desired level of quality as indicated by the metrics for that selection. The selection is applied through thresholds or limits on the generated metrics (see table 2.6 for examples). In applying these thresholds to each of the variables in a study an overall statement of quality can be made for that study.

The visualisation tool in section 5 shows how these thresholds can be determined and applied.

Variable	Operation	Data quality statement (metric)
HbA1c	Completeness through counts	Include practices for which >95 (% of patients with diabetes having at least 2 values, by practice)
Time of diagnosis of diabetes	Completeness through counts	Include practices for which >98 (% of patients with diabetes having a diagnosis time, by practice)
Systolic blood pressure	Accuracy through enumeration	Include practices for which >95 (% of patients with diabetes having systolic measurements to within 2 mmHg, by practice)
Diagnosis of diabetes	Correctness	

Table 6 The application of thresholds to each metric provides a specification of data quality for a study.

Notice that when data is delivered from selected practices, further filters can be applied by the researchers to sharpen the quality further. For example, it may be that the time of diagnosis is mandatory, but that does not mean we set the thresholds in the table above to 100%. We only need the patient records to conform to such stringent criteria, the practices we select less so. A

balance is sought between the chosen thresholds and the quantity of data required to power a study.

3.11 Tasks for determining data quality

The final conceptual model for the data quality assessment is shown in Figure 5

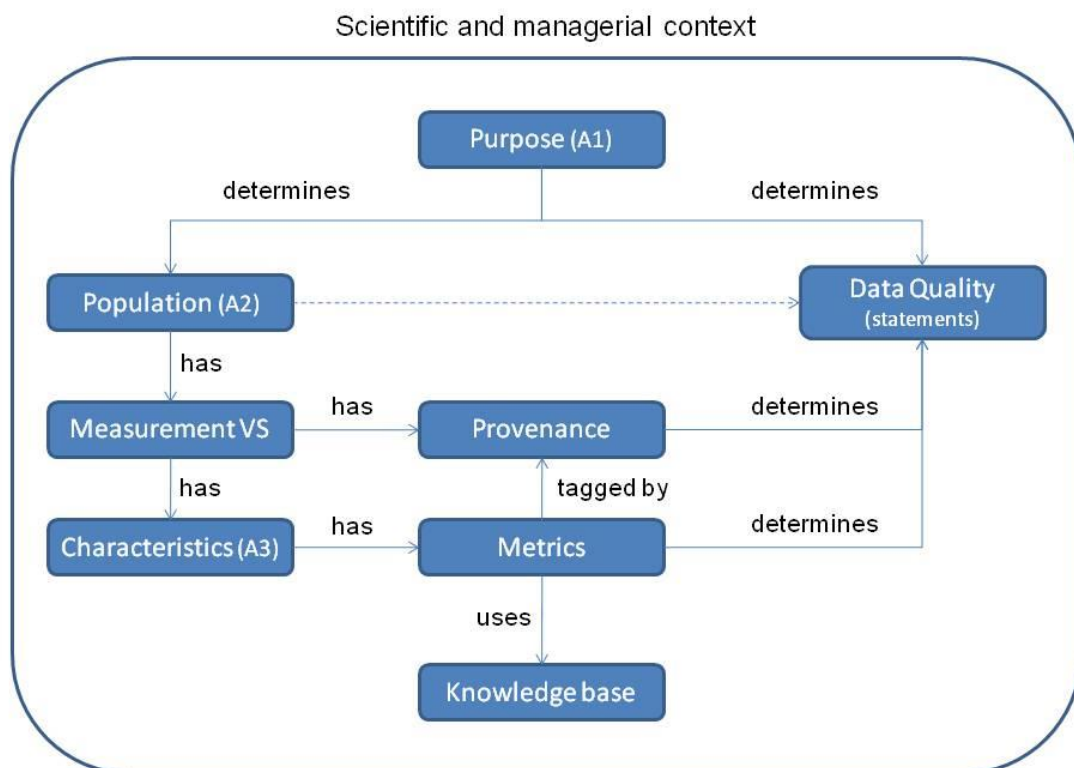


Figure 5. The data quality model

When determining data quality for a variable of interest the following tasks should be performed in order:

- Researcher Define the purpose by specifying formally (using CDIM concepts and reference terminology) the variables of interest (inclusion/exclusion criteria, outcomes, covariates) and the population to which they apply

- Source Define an algorithm for determining the population
- Source Extract the measurement value sets for all the variables for that population
- Source Calculate the agreed metrics using counts or enumerations expressed against known quality determinants
- Source Provide the metrics to the visualisation tool for use by the research community
- Researcher Visualise the metrics against quality determinants and interpret the outputs in terms of available contextual provenance information
- Researcher Set thresholds to select data by any particular determinant
- Source Respond to workbench queries filtered by the selection.

4 Determining Data Quality for TRANSFoRm Diabetes Use Case

In this section we will apply the framework described in section 3 to the TRANSFoRm diabetes use case for the two databases involved in TRANSFoRm: NPCD and GPRD. We will describe the databases, the purpose and an algorithm for determining the population. For all variables the measurement value set is extracted and described in a formal way. Quality metrics are formulated and calculated on the NPCD and GPRD. Finally we will visualise the metrics and develop quality statements based on the metrics, provenance and knowledge base. The TRANSFoRm GORD use case is described in Appendix C.

4.1 Databases

NIVEL Primary care database (NPCD) is funded by The Dutch Ministry of Health, Welfare and Sport, and covers a population of over 1 million patients. NPCD is the successor of LINH (Netherlands Information Network of General practice (36)). It is based on routinely recorded electronic health records from approximately 300 general practices (2012). It contains longitudinal information on claims, prescriptions, referrals, diagnostic test, physiological measurements, contacts and diagnoses. NPCD data can be linked to other datasets using pseudonyms.

The General practice research database, managed by the Secretary of State for Health, United Kingdom (UK) contains 12 million patients overall and 4 million active patient records drawn from approximately 650 primary care practices in the UK (2012). The database has clinical and prescription data.

These databases are summarized in Table 7. More information about coding and classification systems can be found in Appendix A and B.

Databases	NIVEL Primary care database	General Practice Research Database
Size of database		
Country	Netherlands	England
Number of practices	300 GP practices	650 GP practices
Number of patients	1 million	12 million
Number of EHR brands	7 (Medicom, Omnihis, Promedico ASP, Promedico VDF, MicroHIS, Mira)	1 (Vision)

Frequency of data upload	Twice a year	Every month
Content		
Diagnoses/symptoms	ICPC-1	Read codes v2
Diagnostic tests, lab results	NHG codes	Read codes v2
Prescriptions	ATC	Multilex
Claims	NZa codes	
Procedures	NZa codes	
Data available from	1995-2012	1987-2012
Health system characteristics		
Health service characteristics	Managed competition	National Health Service
Basis for reimbursement	Mix of capitation and fee for service	Capitation and quality of care
List system	Yes	Yes
Other		
Record linkage possibilities	Yes	Yes

Table 7 Basic characteristics of NIVEL Primary Care Database and General Practice Research database.

4.2 Purpose and population

The data model's first axiom is "**A1**: *Data quality* can only be determined within the context of a specified *purpose* and related *knowledge*, and that purpose must be explicitly stated (be well-specified)".

The aim of the TRANSFoRm diabetes use case is to create a database of patients with type-2 diabetes containing genetic and phenotypic information compiled from genetic and primary care data sources (repositories). Linkage should be foreseen at the patient level. The purpose of this new database is to answer the following questions.

1. Are well-selected single nucleotide polymorphisms (SNPs) in patients with type-2 diabetes associated with the development of type-2 diabetes related complications as coronary artery disease and retinopathy in a sample of at least 10.000 subjects from the European Union population? **(RQ1)**
2. Are well-selected single nucleotide polymorphisms (SNPs) in patients with type-2 diabetes associated with variations in drug response to oral antidiabetics? **(RQ2)**
3. Which genetic markers (SNPs) are associated with type-2 diabetes related complications such as diabetic nephropathy? **(RQ 3)**

A case control study design for all three questions is foreseen. Only for RQ1, a cohort study design is optional, if enough historical data will be available.

The second axiom states: “ **A2**: A well-specified *purpose* must determine a *population* of interest.”

The purpose of the use case implies that we first need to identify patients with type-2 diabetes. One would think this could be easily done using an ICPC code (in the Dutch example) or a set of Read codes in the UK example. This, however, appeared not to be as simple as that.

In NPCD non-temporary patients, aged 18 or over, with type-2 diabetes are selected with a combination of data concerning diagnosis, prescriptions, claims codes and diagnostic test codes. Patients with type-2 diabetes have:

1. ICPC code T90 (Diabetes) in their record
2. OR have been prescribed blood glucose lowering drugs
3. OR have a code for diabetes counseling in their record
4. OR a code that states that the GP is the main responsible for the Diabetes treatment of the patient.

For GPRD the patients with type-2 diabetes were selected using, first, a string matching approach to find Read codes mentioning diabetes and its type. These codes are manually inspected and the program is adapted to exclude codes such as fh: diabetes (fh= family history), gestational diabetes, diabetes monitoring and other similar ambiguous codes. The result is a list of Read codes that is assumed to indicate Type-2 or Type-1 patients. This list is then used to find these patients in the GPRDs database.

The clinical, referral and test records are searched for a code in the diabetes code list and the first event date that any such code is mentioned is recorded as the index date. Each patient is assigned a category - which is the minimum category in the above list. So, for example if a patient has a code for category 1 they are classified as category 1, even if they also have a code for type-2 (cat. 3 or 4) or a general code. A patient with only a code for category 6 would be classified as category 6, etc.

1. Type-1 with a diagnostic (C) code
2. Type-1 mentioned but no Type-1 diagnostic code
3. Type-2 diagnostic code
4. Type-2 mentioned but no Type-2 diagnostic code
5. General diagnostic code – type not specified
6. General non-diagnostic code – type not specified

For the TRANSFoRm diabetes use case only the categories 3 and 4 were regarded as patients with type-2 diabetes.

4.3 Measurement Value Set (MVS)

The Measurement Value Set (MVS) is the set of values along which the quality of the data pertaining to one variable is actually measured. The data elements of interest, according to the researchers involved in the TRANSFoRm diabetes use case are shown in Table 3.2

Entity	CDIM concept	Classification System (LOINC/ICPC)
Demographics		
Age/Year of Birth	Human birth instant	
Gender	Biological sex	
Race/Ethnicity		
Diagnoses		
Time of Diagnoses	Diagnostic conclusion instant	
Diagnose: Type-2 Diabetes	Diagnostic_ conclusion	ICPC codes: T90 (T90.2?) SNOMED-CT: 44054006 ICD-10: non-insulin dependent diabetes mellitus (obsolete): E11
Diabetic Retinopathy	Diagnostic_ conclusion	ICPC: F83.01 ICD-10: E11.3 (E11.3A, E11.3B, E11.3C differentiates retinopathy from other eye complications)
Diabetes Nephropathy	Diagnostic_ Conclusion	ICPC: U99.1 ICD10: E11.2 E11.2A and E11.2B differentiates nephropathy from other kidney complications
Family history of diabetes	Diagnostic_ Conclusion	
Life Style		
Physical activity habits/sedentarism	Clinical history	
Smoking status	Clinical history	LOINC: 62554-1
Diagnostic test results		
Hba1c (values and dates)	Scalar measurement datum laboratory	LOINC: 59261-8 LOINC: 4548-4

	finding	
Blood glucose (values and dates)(fasting)	Scalar measurement datum laboratory finding	LOINC: 1556-0 LOINC: 14771-0 LOINC: 1558-6
Serum creatinine	scalar measurement datum laboratory finding	LOINC: 2160-0 LOINC: 14682-9
BP Diastolic (values and dates)	Diastolic blood pressure measurement datum	LOINC: 8462-4
BP Systolic (values and dates)	Systolic blood pressure measurement datum	LOINC: 8480-6
Measurements		
Weight	Human mass measurement datum	LOINC: 3141-9
Height	Human height measurement datum	LOINC: 8302-2
BMI	N/Av	LOINC: 39156-5

Table 8 Data elements of interest for the Diabetes Use Case

4.4 Metrics

The third axiom states that: *“The measurement value set is fully characterized by specifying its completeness, accuracy and correctness. No further characterisations are necessary.”*

For every metric that is computed in this paragraph we acknowledge that it is not possible to check the results of each practice in the real world. As stated in paragraph 3.7 we are not interested in any one % value, i.e. the value for a particular practice in a particular year. Instead, we are interested in how that % value varies by practice and time – the quality determinants - which does not require us to know the truth. The variation we see leads us to believe something about whether we should use the group's data or not. Any thresholds put on the computed metrics are dependant on whether the reasearcher is interested in the metrics and has enough data to power the study.

One arbitrary entity is selected out of four categories from Table 8: Diagnoses, Comorbidity, Diagnostic test results and Measurements, and are fully characterized by specifying its completeness, accuracy and correctness.

Other categories and their entities are calculated and presented in the data quality webtool.

Character	Metric
Diagnoses: Type-2 Diabetes	
Completeness	The proportion of the practice population with a diabetes code, by practice code for 2009, 2010 and 2011.
Correctness	% Of patients with a Metformin prescription and a diabetic code recorded
Accuracy	% Of patients with a diabetes type-2 code of all the known patients with diabetes in the practice population (denominator is known diabetes patients)
Consistency	Above metrics are profiled over time.
Comorbidities: Diabetic Retinopathy	
Completeness	% Patients with diabetic retinopathy of all the known patients with diabetes in the practice population in 2009, 2010 and 2011 (denominator is known patients with diabetes)
Measurements: BMI	
Completeness	% Of patients with diabetes for whom a BMI is recorded in 2009, 2010 and 2011
Correctness	% Of all BMI values between 12-50 kg/m ² , calculated per year
Accuracy	% Of all correct BMI values that is accurately recorded in decimal
Consistent	Is the BMI measurement consistent over the years?
Diagnostic test results: HbA1c	
Completeness	% Of patients of which a HbA1c is recorded (denominator is known patients with diabetes)
Correctness	% Of HbA1c value between 0- 300 mmol/mol? Or 0-100%, if recorded in percentage.
Accuracy	% Of all correct HbA1c values that is accurately recorded in decimal
Consistent	Is the HbA1c measurement consistent over the years?

Table 9 Quality Metrics for Diabetes Use Case

4.5 Visualising the Quality metrics and Quality Statements

Inter-practice variation in metrics and results of internal correlations are used as measures of data quality. Thresholds are judged and set to select good quality practices. The percentages in the quality statements are fictional; they are dependent on the quality the researcher requires for his research. To what degree an entity needs to be complete, correct or accurate can be set by the researcher. A balance needs to be sought between the chosen thresholds and the quantity of data required to power the study.

The metrics described in the last paragraph are visualized and quality statements are given for every metric.

Diagnoses: Type-2 Diabetes

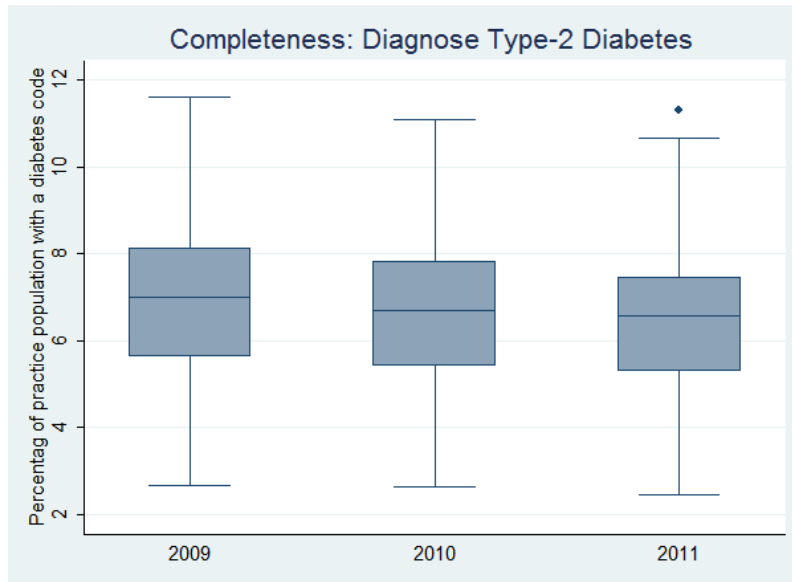


Figure 6 Percentage of non-temporary patients aged 18+ with diabetes code (ever in their record) in 2009, 2010, and 2011 in the NPCD

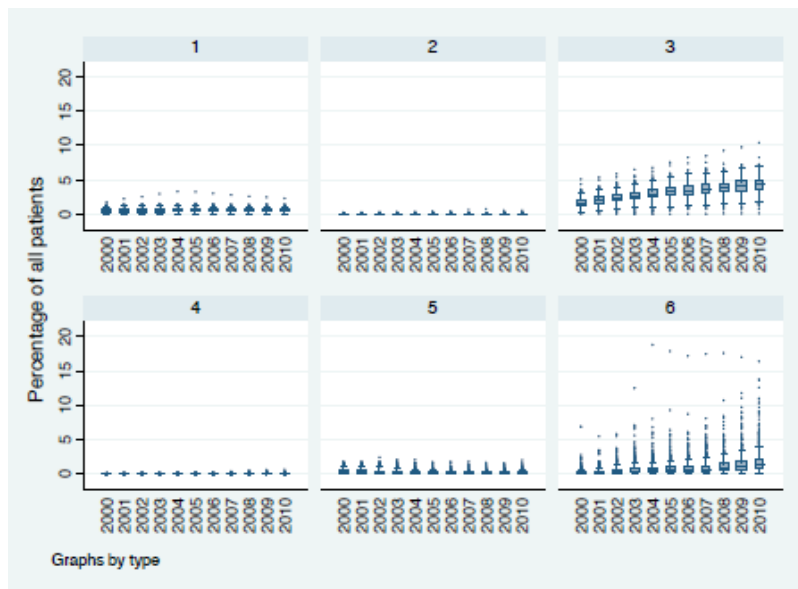


Figure 7 Percentage of registered patients aged 18+ with a diabetes code (in past or present year) and registered and alive in each year 2000-2010, by patient's code category in GPRD

Data entity	Metric	Quality Statements
Diagnose	Completeness through distribution	Include practices for which >5% of the practice population have a diabetes code recorded
Diagnose	Correctness through cross-checking	Include practices for which >95% of patients with a metformin prescription also have a diabetes diagnoses recorded
Diagnose	Accuracy	Include practices for which >90% of the

patients have a type two diagnoses code

Comorbidities: Diabetic Retinopathy

The completeness of diabetic retinopathy is calculated through counts and profiled by practices over years (quality determinant). For practices with 0% of the diabetes population with a diabetic retinopathy code one can dispute whether the recording of comorbidity, in this case diabetic retinopathy, is complete. By putting threshold on the outlier practices, the researcher can select practices with more complete data.

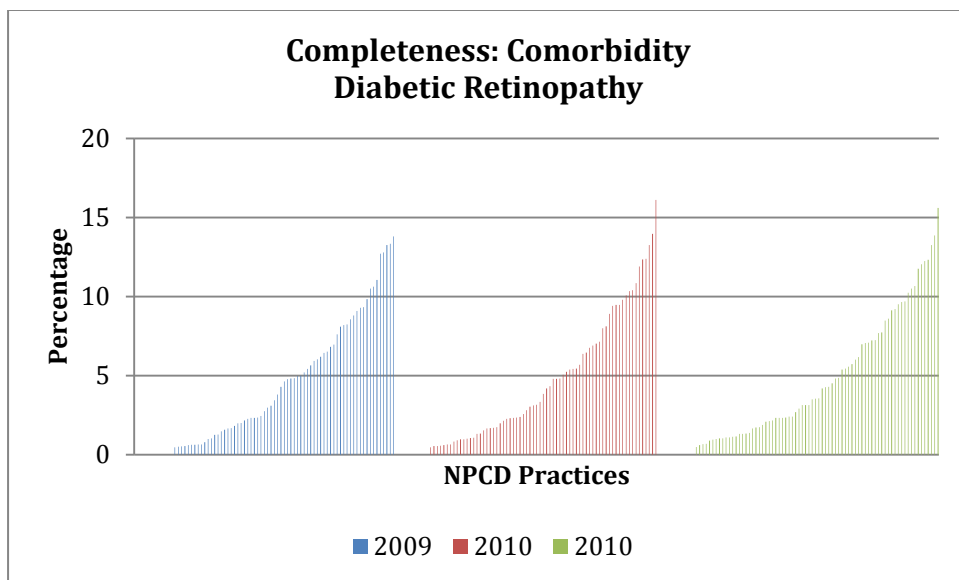


Figure 8 The proportion of diabetic Retinopathy in the know diabetic population in 2009, 2010 and 2011 (NPCD)

Diagnostic lab result: HbA1c

The completeness of HbA1c recording is getting better every year in the NPCD. The correctness is stable, and the accuracy of the HbA1c recording becomes successively lower every year. We can understand the graphs with the contextual knowledge we have. Which is that the NPCD was able to extract more diagnostic test results over the years, which means that different EHR-vendors were included. This made the data more complete, but with more data coming from different EHR-vendors, the format of the incoming data also varied. The new EHR-vendors didn't deliver more accurate data, so consequently the accuracy went down, and eventhough data was more complete. The data quality determinant for this metric is time (the years in which the HbA1c measurement was recorded).

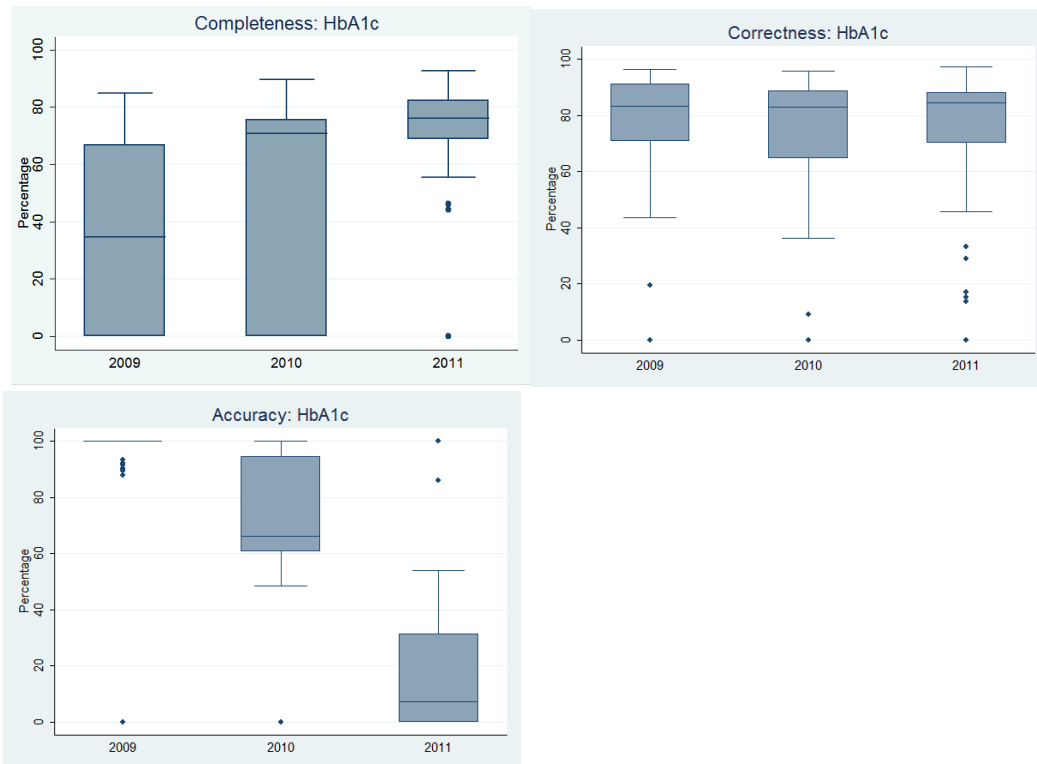


Figure 9 HbA1c metrics between years and within practices is calculated for 2009, 2010 and 2011 (NPCD). Vertical bars show upper and lower limits of practice completeness, correctness and accuracy, 25% and 75% quartiles and medians.

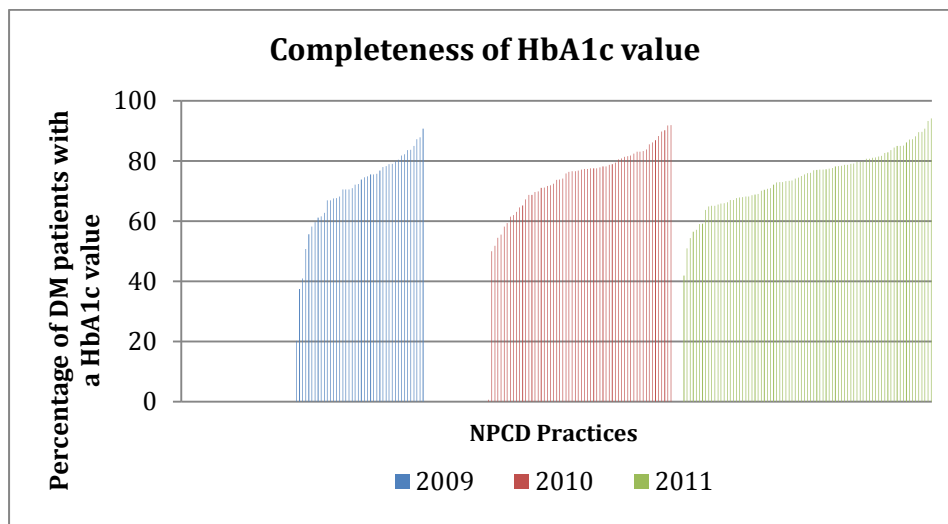


Figure 10 the proportion of patients with type-2 diabetes within a given practice having at least one diagnostic lab result (HbA1c) in 2009, 2010 and 2011 is computed. Practice is here the quality determinant.

Data	Metric	Quality Statements
------	--------	--------------------

entity		
HbA1c	Completeness through counts	Include practices for which >90% of diabetic patients have a value
HbA1c	Correctness through enumeration	Include practices for which >70% of all HbA1c values are between 0-300 mmol/mol Or 0-100%.
HbA1c	Accuracy through enumeration	Include practices for which >10% of HbA1c values are accurately written in decimal

Measurement: BMI

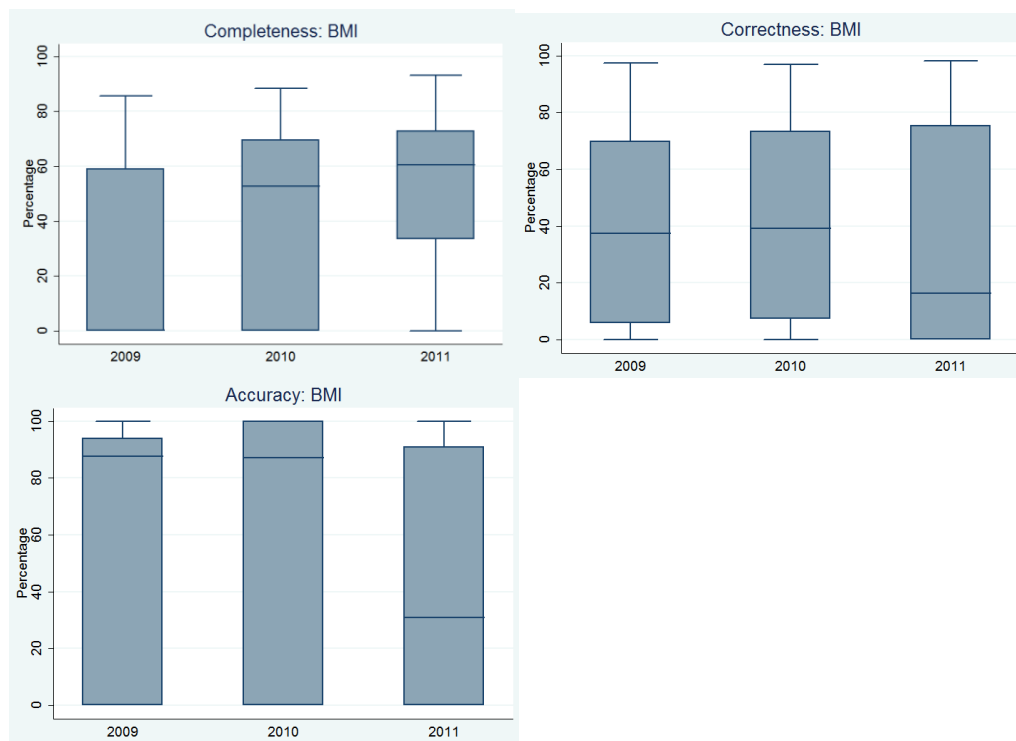


Figure 11 BMI metric between years and within practices is calculated for 2009, 2010 and 2011 in NPCD. Vertical bars show upper and lower limits of practice completeness, correctness and accuracy, 25% and 75% quartiles and medians.

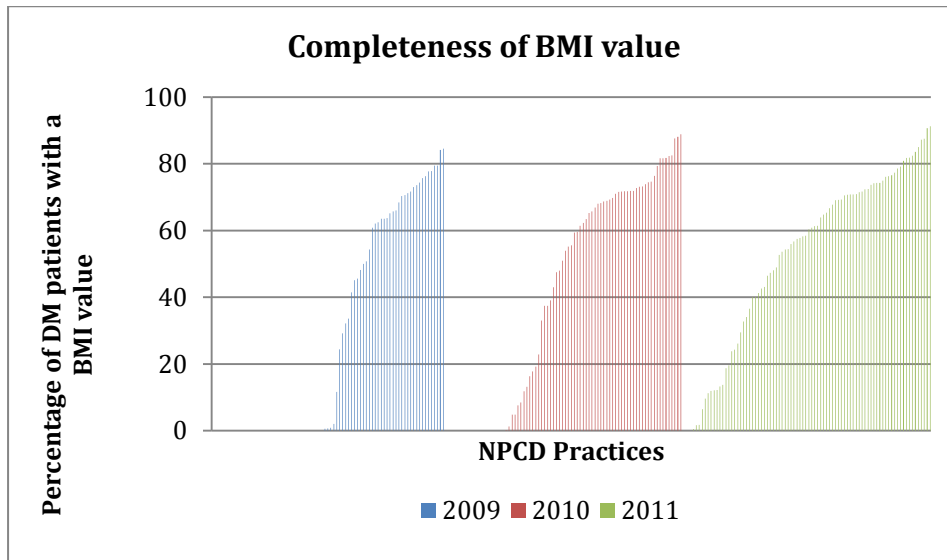


Figure 12 The proportion of patients with type-2 diabetes within a given practice having at least one BMI measurement in 2009, 2010 and 2011 is computed.

Data entity	Metric	Quality Statements
BMI	Completeness through counts	Include practices for which >80% of diabetic patients have a value
BMI	Correctness through enumeration	Include practices for which >30% of all BMI values are between 15-50 kg/m2
BMI	Accuracy through enumeration	Include practices for which >10% of BMI values are accurately written in decimal

The quality metrics compared

For the quality metric completeness NPCD and GPRD could be compared. You can see that the researcher should keep in mind that the thresholds should not only be set looking at one database, because if set to too stringent criteria, a lot of practices will be lost. A balance must be sought between the chosen thresholds and the quantity of data required.

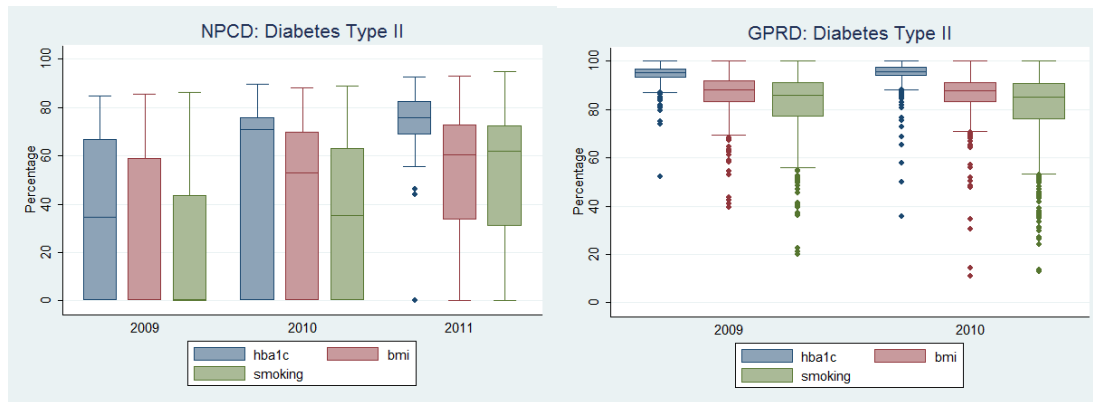


Figure 13 the proportion of patients with type-2 diabetes within a given practice having at least one measurement of a given type within a given year.

4.6 Discussion

In this section we used the data quality framework described in section 3 to find appropriate quality metrics for the TRANSFoRm Diabetes use case. This use case was our focus for developing the quality metrics. We agree that more quality metrics can be defined if the researcher feels necessary.

Following the framework we first set out to determine the population in two databases, adhering to the way this is usually done in each respective database. One of the conclusions here is that identifying the population of patients with type-2 diabetes cannot be based on the same query in both countries. In the Dutch database patients with type-2 diabetes are selected on different criteria as compared to GPRD.

While working on this deliverable it was not possible to use the Clinical Data Integration Model (CDIM), which was simultaneously developed in WP6. This resulted in that we were not able to test the mapping of different classification systems. A lot of discussion and comparing was done, in order to find the right codes.

5 TRANSFoRm Data Quality Webtool

In section 3 we discussed a framework for data quality that starts with formulating a purpose. The purpose determines the population and the data elements needed, and thereby the data quality metrics. This implies that there is no universal set of data quality metrics. Metrics always will look different, dependent on the purpose. Examples of data quality metrics for diabetes use case in TRANSFoRm are described in section 4 and will not be discussed here. In the present section we will only discuss the webtool itself. For the current section the question is how to incorporate such quality metrics into a webtool that will enable researchers with a specific research purpose to select practices with ‘fit for purpose’ data.

A demo version of the webtool is available at:

- <http://nvl007.nivel.nl/apps/transform/>
- Email: transform@project.eu
- Password: TR@nsform

5.1 Requirements for the Data Quality Webtool

The design and implementation of the Webtool was based on the requirements-driven approach. We used the use cases provided by deliverable D1.1. A participatory design session with the authors of the reports in the beginning of the development process defined the requirements for the webtool.

1. The system should be **flexible**. Other data quality metrics must be incorporatable, dependent on the purpose.
2. Database **independent**. It should be possible to report on data quality metrics, irrespective of the underlying structure of a given database.
3. **Authorize** data extraction. Only researchers involved with the TRANSFoRm project should be able to login.
4. **Select** practices which are ‘fit for purpose’.
5. **Extract** information for the selected practices and metrics for use within other TRANSFoRm tools.
6. The researcher should be able to **save** the queries.
7. It should have the possibility for researchers to comment and give **feedback** to data sources.

Given these requirements we developed a prototype of a Data Quality Webtool which is available for two databases: NPCD and GPRD.

5.2 Architecture of the Data Quality Webtool

This paragraph gives a brief description of the system architecture and its functional components, which form the context for the design and implementation of the Data Quality Webtool, as defined by the requirements given in the last paragraph.

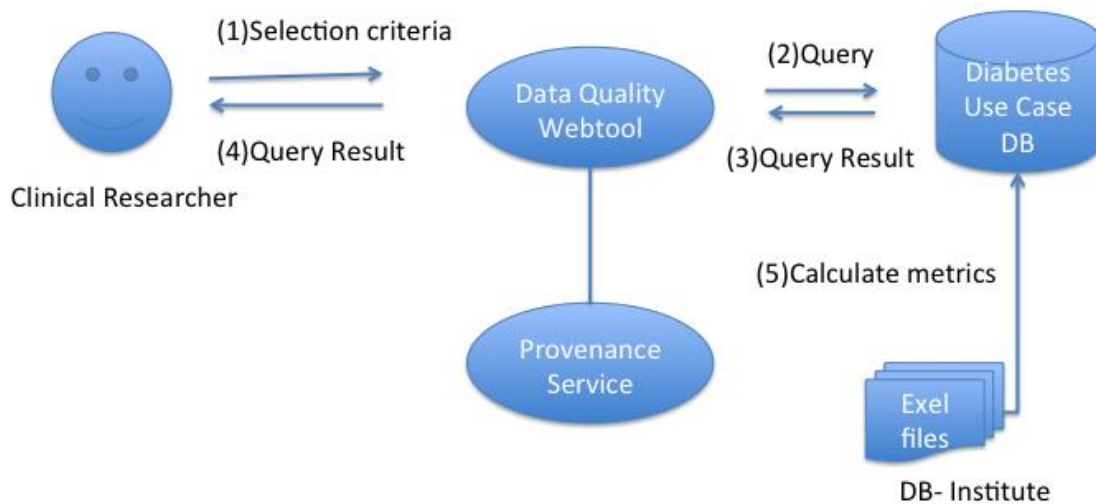


Figure 14 Conceptual Architecture

An essential part of this workflow focuses on supporting clinical researchers in finding appropriate thresholds for practices with good quality data to include in research studies. These thresholds can differ for different metrics a balance is sought between the chosen thresholds and the quantity of data required to power a study.

Calculate Metrics

Input for the webtool takes the form of a predefined use case specific table (Microsoft Excel) with practice IDs in the rows and quality metrics in the columns. This table can be generated by the datasources or through the TRANSFoRm platform. An example of a table is given below; a template is available in the webtool.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	registry	annum	practice_ID	totalNofpatients	Nhba1c	Nbmi	Nbp	NdiasBP	Nsysbp	Nsmoking	Nlength	Nweight	Nglucose	Nalcohol	Ndrugs	pct_hba1c	pct_bmi
2	NIVEL	2009	8979	102.00	63.00	59.00	77.00	77.00	77.00	34.00	82.00	63.00	72.00	42.00	0.00	61.76	57.84
3	NIVEL	2009	8936	481.00	164.00	295.00	333.00	329.00	332.00	112.00	338.00	309.00	329.00	38.00	0.00	34.10	61.33
4	NIVEL	2009	8855	362.00	247.00	228.00	306.00	306.00	306.00	166.00	250.00	144.00	289.00	121.00	0.00	68.23	62.98
5	NIVEL	2009	8459	153.00	108.00	99.00	115.00	115.00	115.00	100.00	128.00	103.00	122.00	24.00	0.00	70.59	64.71
6	NIVEL	2009	8383	72.00	14.00	0.00	3.00	3.00	3.00	9.00	4.00	0.00	61.00	5.00	0.00	19.44	0.00
7	NIVEL	2009	8692	172.00	0.00	0.00	0.00	0.00	0.00	0.00	53.00	0.00	0.00	0.00	0.00	0.00	0.00
8	NIVEL	2009	8415	147.00	0.00	0.00	0.00	0.00	0.00	0.00	83.00	0.00	0.00	0.00	0.00	0.00	0.00
9	NIVEL	2009	8699	157.00	110.00	75.00	103.00	102.00	102.00	52.00	107.00	93.00	112.00	52.00	0.00	70.06	47.77
10	NIVEL	2009	8449	77.00	27.00	45.00	52.00	52.00	52.00	4.00	29.00	50.00	55.00	0.00	0.00	35.06	58.44
11	NIVEL	2009	8187	183.00	0.00	0.00	0.00	0.00	0.00	0.00	118.00	0.00	0.00	0.00	0.00	0.00	0.00
12	NIVEL	2009	8331	130.00	0.00	0.00	0.00	0.00	0.00	0.00	8.00	0.00	0.00	0.00	0.00	0.00	0.00
13	NIVEL	2009	8467	163.00	0.00	0.00	0.00	0.00	0.00	0.00	8.00	0.00	0.00	0.00	0.00	0.00	0.00
14	NIVEL	2009	8219	8.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
15	NIVEL	2009	8883	132.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
16	NIVEL	2009	8958	190.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
17	NIVEL	2009	8948	186.00	0.00	0.00	0.00	0.00	0.00	0.00	17.00	0.00	0.00	0.00	0.00	0.00	0.00

Figure 15 Quality metrics table

For any given purpose (use case), a set of quality metrics will have to be defined in general terms. These terms need to be formulated in such a way that they can be understood and operationalized by the query workbench or any other entity that will provide the table. This can also be the data source itself. The exact parameterization of the queries underlying the metrics will vary between databases.

Provenance Framework

The TRANSFoRm provenance framework (Deliverable 3.1) keeps track of the origin of various data and its evolution in different stages of the clinical studies. Making TRANSFoRm systems provenance aware, one can enable the investigation of data sources and the services that produced a particular output, together with the individuals who instigated the requests and received the outputs. In such a way, user behavior and data manipulation can be audited, to assess that correct decisions were made and appropriate procedures were followed.

API's

To communicate with other software tools in the TRANSFoRm project, such as the Query Workbench (Work package 5.3), API's can be built. The output of the data quality webtool, for example the database- and practice identifiers, will be available as input for other tools.

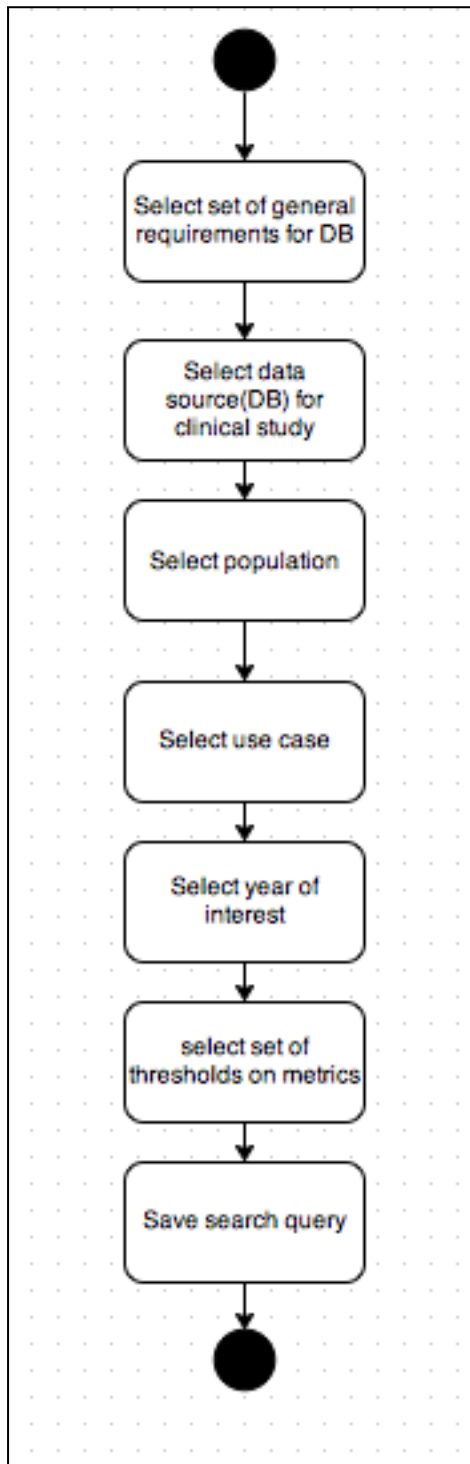


Figure 16 Workflow for selecting practices

5.3 User Interface Design and Workflow of the Data Quality Webtool

The data quality webtool UI design and workflow are described in this paragraph, based on the functional and user requirements identified in paragraph 5.1. The UI workflow is described textually with the help of printscreens of the data quality webtool and with the use of UML activity diagram.

All user interactions with the data quality webtool are in the form of standard web application interactions, such as mouse clicks and simple text input.

Figure 16 shows the overall process of selecting and saving practices that are ‘fit for the purpose’.

The user needs authorization to use the data quality webtool. A login and password is provided on request.

E-mail:

Password:

Figure 17 Authorisation page

After logging in the user can find information about the available databases by clicking on a map. Background information regarding the available databases in that country is shown including contact details of the database owner.

Available registries	
Registry:	GPRD
Name of the registry:	Clinical Practice Research Datalink
Host institution:	Medicines and Healthcare Regulatory Agency (MHRA)
Host contact e-mail:	admin@gprd.com
Host contact phone:	+44 (0) 20 7084 2383
Controlling institution:	MHRA
Controller contact:	John Parkinson
Controller e-mail:	john.parkinson@mhra.gsi.gov.uk
Geographical coverage	UK
Legal jurisdiction	England & Wales, Scotland
language	English
Type of system	General Practice Repository
DBMS	Proprietary
Classification	Read codes
Publication	http://www.cprd.com
Data source	GPIIS
period	1993-now
Number of practices	600
Number of patients	54000000




Figure 18 Background information about the available databases

As a second step, the user can tick out a number of general requirements for any database of which the user wants to have more information about. For example the user may only be interested in databases that can be linked to genetic data.

Select database

<input type="checkbox"/> Patient consent	<input checked="" type="checkbox"/> Linkable to genetic data
<input type="checkbox"/> Contain Physical Examination data	<input type="checkbox"/> Linkable to a cancer registry
<input type="checkbox"/> Contain Lifestyle data	<input type="checkbox"/> Linkable to a drug registry
<input type="checkbox"/> Contain Medication data	<input type="checkbox"/> Linkable to a hospital registry
<input type="checkbox"/> Contain Lab results	<input type="checkbox"/> Linkable to a population registry
<input type="checkbox"/> Contain genetic markers	

Figure 19 Selection of general requirements for the databases

These requirements determine the choice of available databases in the next screen. In this example, only databases will be shown that are linkable to genetic data. In the next step a database can be selected.

Figure 20 Selection of database

As a fourth step, the user gets information about the completeness of some basic data elements that are present in the database in terms of the four aspects of data quality described in section 2: correctness, completeness, accuracy and consistency (if available). The aim of this section of the data quality webtool is to give the user general information about what is in the database, irrespective of any use case.

Database information			
Completeness	Correctness	Accuracy	Consistency
episode		episode	
prescription			
contraindication			
patient	patient		
results			
claim			
referral			
morbidity			

Figure 21 General information about the selected database

As a fifth step, practices can be selected that satisfy the needs for a certain study. In the prototype this is done for the diabetes use case and the GORD use case (see deliverable D1.1). This fifth step starts with defining a use case, a population and a time period.

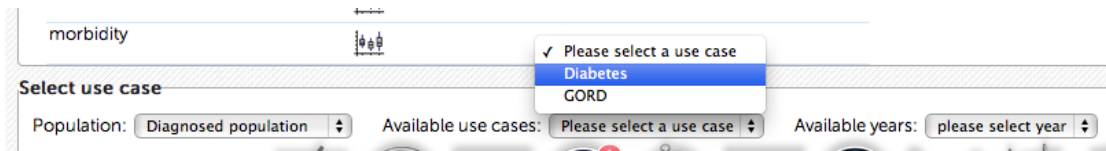


Figure 22 Selection of the use case

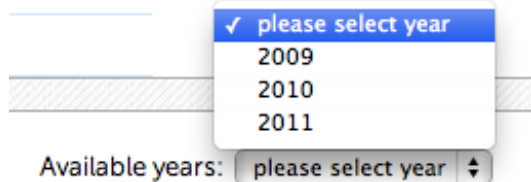


Figure 23 Selection of the year

After that a number of data elements appear that are considered relevant for the study at hand. These data elements are predefined. The slide bars enable researchers to apply quality criteria more or less strictly, and to get informed instantly about the number of available practices and patients therein that satisfy the criteria. At both ends of the slide bars the minimum and maximum values are shown, so that the user is informed about the values between which the sliders can move.

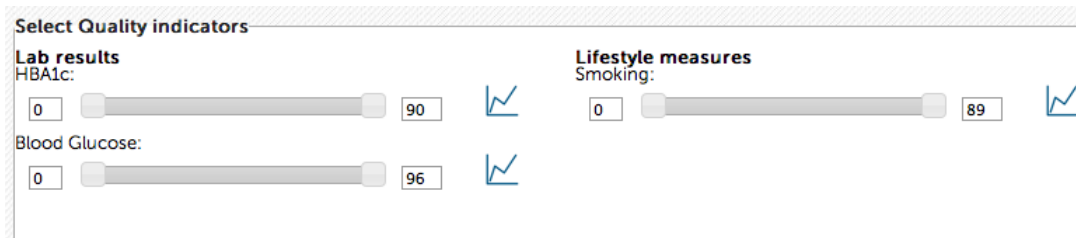


Figure 24 Data elements for the use case

To get a general idea of how the metric is distributed across practices, the total distribution can be shown by clicking the 'graph' icon.

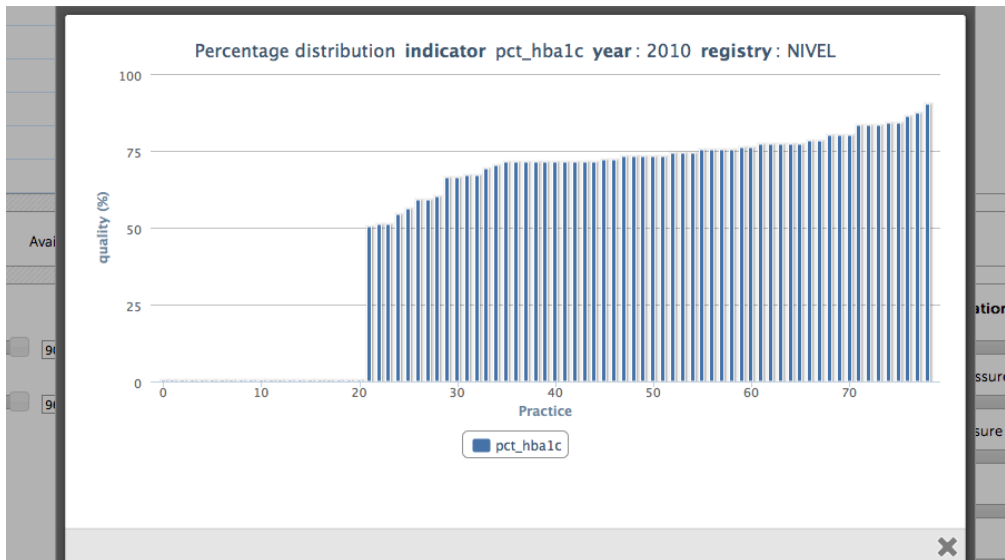


Figure 25 Distribution of HbA1c metrics over practices in NPCD in 2010

At the bottom of the webtool the cumulative results of the selections are shown in terms of practices that satisfy the criteria. Moving one or more of the sliders will result in a decrease or increase in the number of practices that would be involved in the study. The different sliders are cumulative.

Current selection

Show 10 practices

practice_ID	TotalNofpatients	HBA1C	BMI	BP	DiasBl
8007	15	12	10	11	10
8011	194	109	84	138	138
8050	150	101	24	119	119
8052	145	0	0	0	0
8057	163	0	0	0	0
8087	373	288	297	324	324
8111	179	0	0	0	0
8115	198	101	99	135	135
8121	196	140	105	134	0
8125	161	121	118	136	136
Total:	15924	8215	6554	8816	8660

Showing 1 to 10 of 78 practices (filtered from 242 total practices)

Figure 26 Result of the query

Identifier for these practices is a practice ID. A list of practice IDs can be fed back into the TRANSFoRm platform to allow platform queries to be directed at the appropriate practices. The threshold values that were used, as well as the results of the selection process can be saved in an excel file.

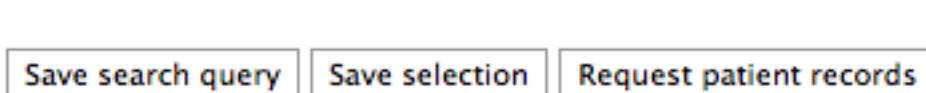


Figure 27 Possibility to save the query or send a request to the database

The prototype provides the possibility for users to give feedback. In the prototype this is done using a free text format (at the top of the webpage).

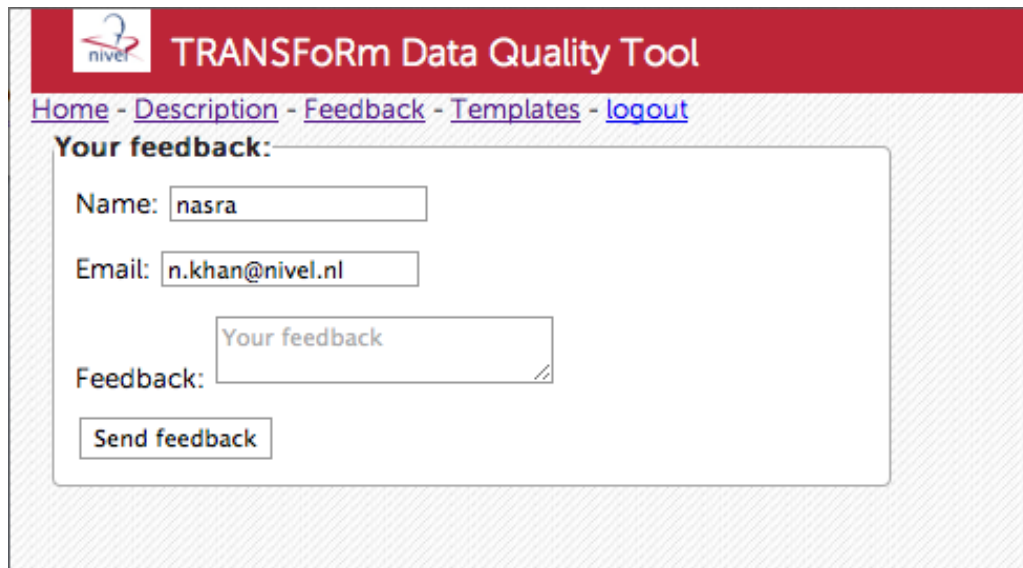


Figure 28 Feedback about the data quality metrics

5.4 Evaluation of the Data Quality Webtool

Subsequent versions of the data quality tool were evaluated by investigators of the two TRANSFoRm uses cases. Comments on previous versions were used to construct new versions to ensure that the data quality itself would be fit for purpose in the two use cases. In addition, however we performed a more formal evaluation using a questionnaire based on the Technology Acceptance Model (TAM).

For the past 30 years, significant developments have been made to explain and predict user acceptance of information technology (40), by evaluating user behavior based on factors of usability, usefulness and technology-based attitudes of individual users. Predominantly, considerable theoretical and empirical support has accumulated in favour of the Technology Acceptance Model – TAM (41) (42). This particular theory has been replicated by numerous researchers, who tested different types of information systems in diverse domains of knowledge, allowing the performance of hypothesis driven evaluation of such systems. Holden and Karsh have recommended to specifically adapting TAM to the healthcare context, with some changes in the constructs of the model (43). These recommendations, together with the most recent evolution of TAM, based on Venkantesh’s the Unified Theory of Acceptance and Use of Technology (UTAUT), can provide useful insights to

usage behavior and intention for newly introduced information technology and tools.

To evaluate the webtool we used a TAM-inspired questionnaire, based on previous work by Algharibi and Arvanitis, on the use of UTAUT in the evaluation of e-trials systems (44). The specific questionnaire investigates factors that affect user related performance and effort expectancy in the use of novel information systems and tools, as well as, the effects of social influence, usability facilitating conditions, technology anxiety and adaption timeline, in order to understand users' behavioural intention in the future use and adoption of e-trials technology. Although, the model in (44) is a hypothesis-driven model, here, for the purposes of the evaluation of the data quality web tool, we use the constructs of the model in a more descriptive fashion, to understand individual factors of usage behavior and intention. A full Technology Acceptance Modelling study will take place in the evaluation and validation phases of the project, where TRANSFoRm tools and services will be fully evaluated as part of clinical scenarios. The factors that we evaluated and are described below correspond to the groups of questions of the data quality web tool questionnaire, available in Appendix D.

Questionnaires were sent to a convenience sample of 30 potential users, representative to typical roles (see below) who were affiliated with TRANSFoRm and/or working with EHR data for research purposes in one of the partnering institutes in TRANSFoRm. A total of 26 questionnaires were sent back; 2 had to be disregarded for being incomplete, therefore $n=24$. For the initial results, a descriptive statistical analysis was conducted on the variables of Individual Factors to depict the participants' demographics. Their ages varied from 20 to 61 years, yet the age group (32-37) had the highest rate at 29.2% in comparison with the other age groups. The male to female ratio was 1:1. The majority of participants were Clinical Researchers (45.8%). Not everyone was experienced in conducting clinical research; 75% had previous experience in clinical research and within those 45.8% had experience ranging between 1 and 6 years. Most participants (62.5%) stated that they were using the DQT System voluntarily, in order to improve their work performance. All questions were based on a 7-point likert scale, with ratings corresponding to 1-strongly disagree, 2-very much disagree, 3-disagree, 4-not sure, 5-agree, 6-very much agree and 7-strongly agree. All questions were previously validated (44) while a Cronbach's Alpha reliability analysis was also conducted to test the consistency of the likert scales. For the purposes of the discussion below, we will provide individual statistics on average ratings (and their standard deviation) from individual questions, while

we will report the overall percentage of respondents for rating 5 to 7, as a positive response.

Figure 29, shows factors relating to individuals’ performance expectancy. According to Davis (42), this describes the “...degree to which a person believes that using a particular system would enhance his or her job performance.”

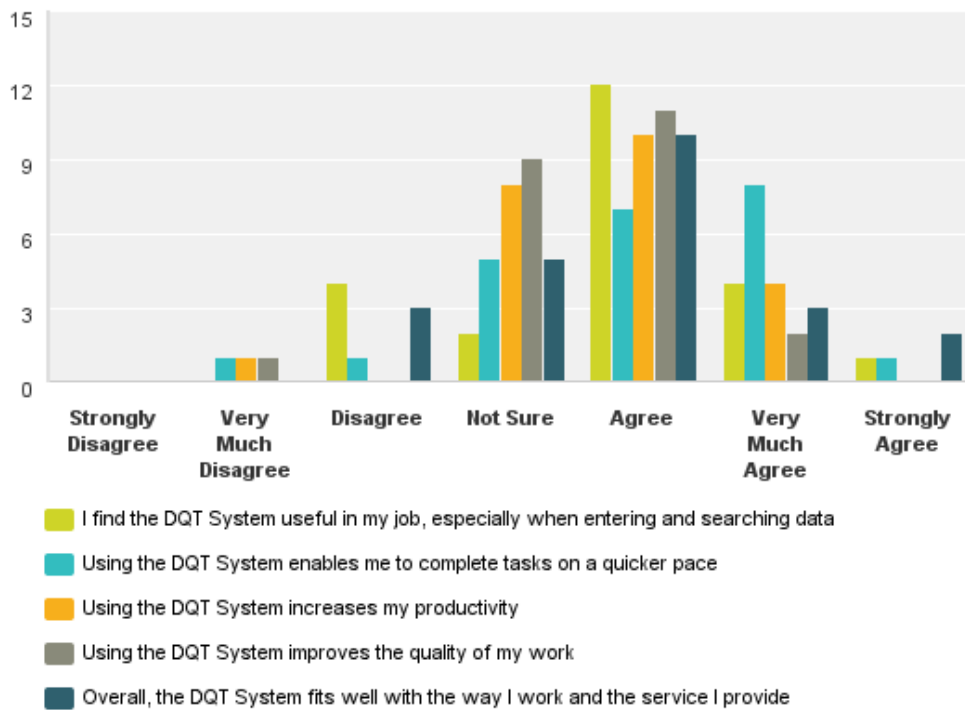


Figure 29 Performance expectancy evaluation

Overall the users expected the tool to improve their work performance. They positively reported that the data quality web tool can enable faster completion of tasks (average rating 5.00 ± 1.17 , with an overall positive response of 70%) while reported their perception of improved productivity (average rating 4.70 ± 0.93 , with an overall positive response of 61%). In particular, a higher positive response was reported in the tasks of entering and searching data (average rating 4.83 ± 1.07 , with an overall positive response of 74%). The overall positive response on how the tool fits in their way of working was 65%.

Figure 30, shows the factors that relate to the data quality tool users’ effort expectancy, which according to Davis (42) describes “the degree to which a person believes that using a particular system would be free of effort”. Here user-friendliness in terms of becoming skillful in using the system, speed of

time of learning to operate the system, and no exertion in the effort for using the system were highly rated, with overall positive responses of 74%, 78% and 78%, respectively (average ratings of 4.87 ± 0.76 , 5.09 ± 1.16 and 5.26 ± 1.05 , respectively). The overall user-friendliness had a more average positive response of 64% (average rating of 4.65 ± 1.39) and users were more moderate in their perception on the facilitation the tool could currently offer to their work (average rating 4.26 ± 1.05 , with an overall positive response of 48%).

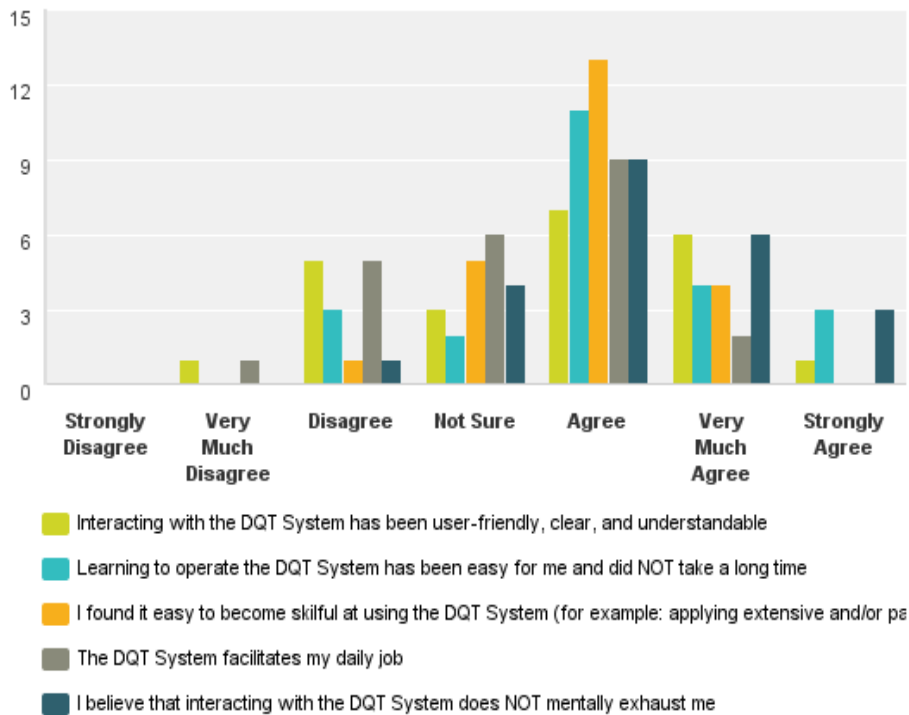


Figure 30 Effort Expectancy evaluation

The influence of the organizational structure, the long term benefits of the tool and the importance of the introduction of such tools in the users’ workplace have been evaluated through the TAM related construct of social influence. This construct, according to Aizen (45), describes the user’s “perceived social pressure to perform or not to perform a behavior” with technology. Here, as shown in figure 31, our users have reported highly positive responses on the importance of use of such software within their organization and the good communication and training on healthcare ICT by their managers and their organization. However, they have shown some uncertainty if their managers/supervisors have shown encouragement in the use of the data quality web tool. In the question/statement “My managers/supervisors

encourage me to use the DQT System” the overall average rating was 3.96 ± 1.15 , with a positive response of 21%, while a response of “not sure” was at the level of 57%.

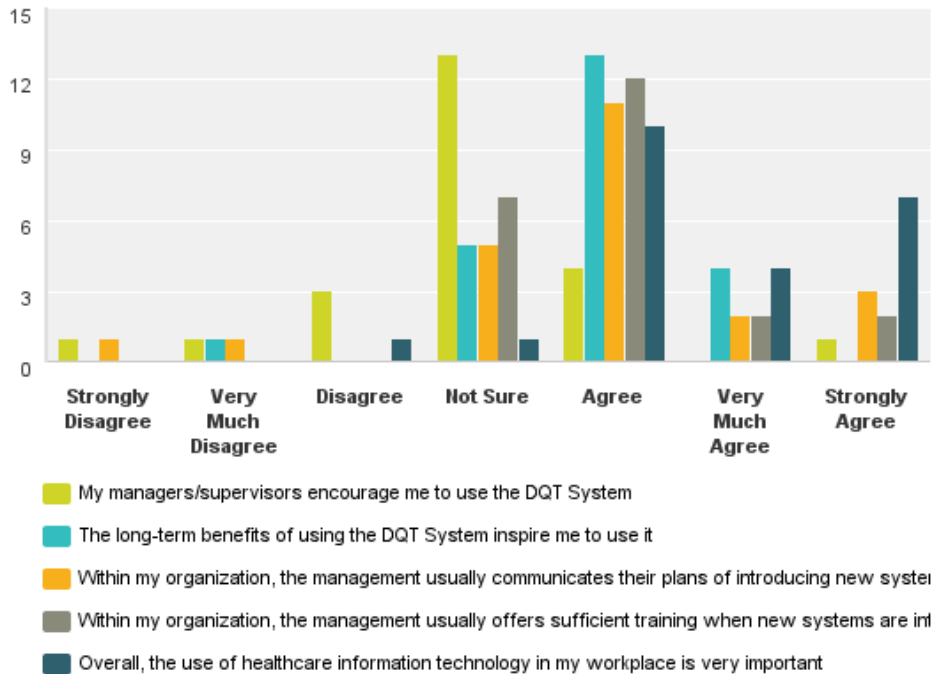


Figure 31 The effect of Social Influence

The steps, which the tool follows, were logical to use, apply and recall. The researchers could imagine using the system to conduct their routine work. These aspects of the system were evaluated with the construct of facilitating factors. Venkanthesh (40) refers to these factors as “objective factors in the environment that observers agree make an act easy to accomplish”. In an Information Systems context Thomson (46) defines this construct as “the provision of support for users of PCs may be one type of facilitating condition that can influence system utilization”. Table 10 and figure 32, summarise the results relating to facilitating factors; users are overall satisfied but they note that the help function/support in the tool needs further improvement.

Facilitating factors	Average Response	STD	Positive response
The training/help option was/is comprehensive enough	4.96	1.02	65%
The use of the DQT System is smooth (no technical errors)	5.39	0.78	92%
The steps followed by the system are logical to use, apply, and recall	5.17	1.07	83%
If I got stuck, a sufficient reference or resource is available to help me out	4.39	0.84	43%
The DQT System is easily accessible, responds to queries promptly and does not disturb my daily routine	5.26	0.92	83%

Table 10 Summary results for the facilitating factors construct of TAM.

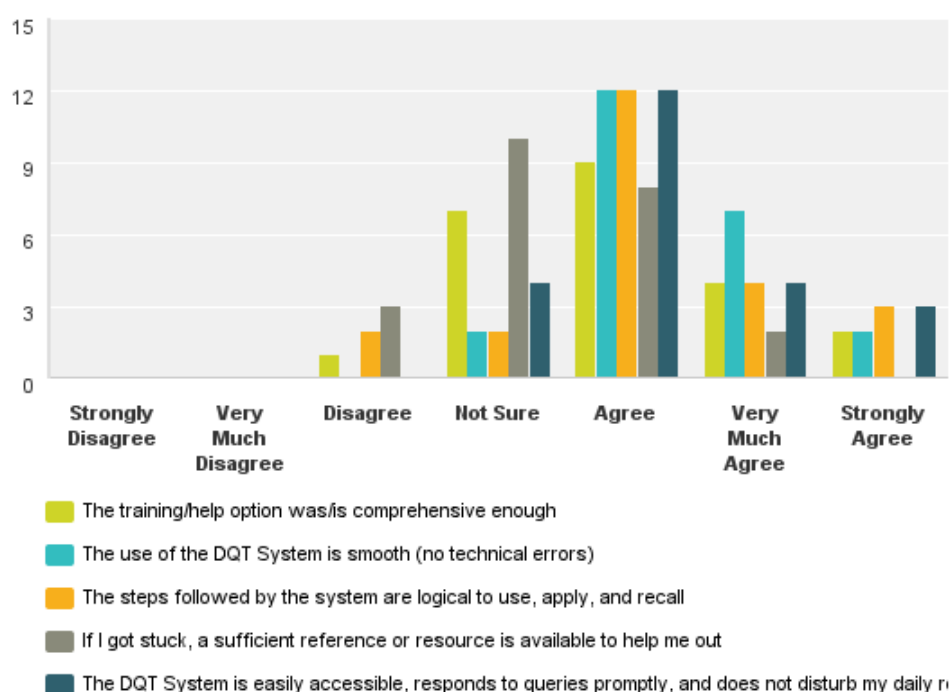


Figure 32 Facilitating Factors

As part of this evaluation, we have identified specific features that the tool offers to the users, which were important to be evaluated. Elements that relate to the login to the system, logical workflows, the visual presentation and manipulation of results, the data quality tool ID and its use within the Query and Data Extraction workbench have been looked by the users. Overall, a good positive response is identified for all elements, although the evaluating users had an average, but still positive response to the use of the practice ID provided by the tool, for its use to the query workbench, An automatic integration of the workbench with this facility, was identified as a much

preferred solution, on the qualitative comments given by the users. Table 11 shows the average responses and overall positive response percentages for the specific features of the tool.

Data Quality Tool - Specific Features			
	Average Response	STD	Positive response
DQT system: the number of steps, to select practices fit for purpose, is reasonable	5.39	0.89	91%
DQT login: logging into the system is quick enough	5.61	0.99	96%
DQT information map: there is enough information about available databases, their country of origin, and owner contact details	5.00	1.21	78%
DQT specific database: for a database of interest, some further general requirements can be selected such as: whether the database is linkable to generic data, to a cancer registry, to drug registry, etc.	5.09	1.12	74%
DQT data quality: the system assesses the completeness of available data in a visual output which is helpful to interpret	5.04	1.22	74%
DQT defining use case: querying data for a certain study using use case, population, and time period is prompt enough	5.39	0.99	96%
DQT specifying use cases: after a use case is defined, further specification can be applied using slide bars on predefined data elements	5.39	1.03	92%
DQT output: the practice ID given for a specified use case is the right output	4.96	1.15	61%

Table 11 Summary results for the evaluation of the specific features of the Data Quality Web Tool.

The researchers were not anxious about using technology. Table 12 and figure 33 show the results that relate to the construct of behavioural intention for the future adoption of the tool and its promotion to other researchers in the field. Overall the users tend to be positive in the adoption of the Data Quality Tool, if allowed enough time to get into the habit of using this software artifact, if it becomes available. They also intend to do more research studies if the system proved to substantially facilitate so.

Behavioural Intention			
	Average Response	STD	Positive response
I intend to learn how to use the DQT System to the best of my abilities	5.22	1.04	78%
I intend to use the DQT System as often as needed to perform my work	5.00	1.13	70%
I intend to promote the DQT System to my colleagues if it became available for broader use	5.43	0.99	83%
I intend to suggest ideas to further enhance the DQT System if I had any	5.43	1.12	78%
I prefer to use the DQT System because data can be better preserved, archived, shared, manipulated, and	5.30	1.11	74%

Table 12 Summary results for the user’s behavioural intention in the future adoption and use of the Data Quality Web Tool.

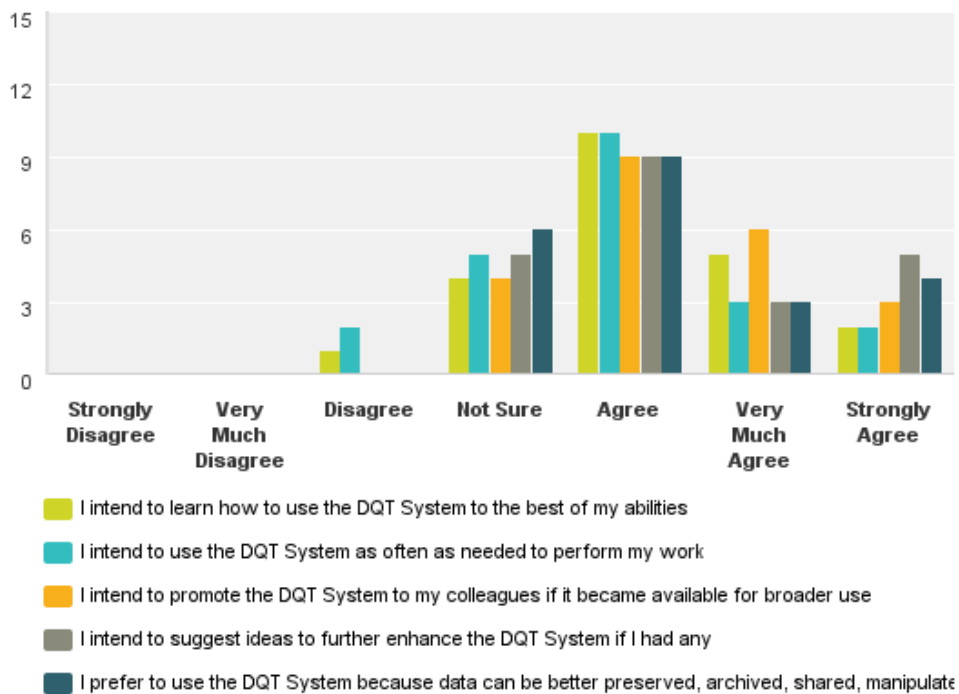


Figure 33 Behavioural Intention Construct for the DQT TAM evaluation

Appendices

Appendix A: NIVEL primary care database: classifications and coding

Diagnosis coding

The International Classification of Primary Care (ICPC) allows for the classification of the patient's reason for encounter (RFE), the problems/diagnosis managed, primary or general health care interventions, and the ordering of the data of the primary care session in an episode of care structure (37;38).

The ICPC structure is bi-axial: the first axis consists of 16 chapters relating to body systems and one chapter on social and societal problems, and the second axis consists of components, by which diagnosis, symptoms, diagnostic and therapeutic interventions and other aspects of the disease episode can be recorded.

General practitioners in the Netherlands use the first version published in 1986, ICPC-1. The Dutch GPs College (NHG) releases the Dutch translation. More than 90% of Dutch GPs are coding the diagnosis with ICPC-1 which forms an integral part of all authorized electronic health record systems in general practice.

The capital letters A-Z indicate the body systems/problem areas. An ICPC code always has an alpha for the chapter, and two digits for the rubric in the component, e.g.: Heartburn: Chapter D(igestive), symptom/complaint component 1: D03 Pneumonia: Chapter R(espiratory), disease component 7: R81.

The ICPC code for diabetes is T90. Subcodes are used to indicate type-1 (T90.01) and type-2 (T90.02). In Dutch general practice the subcodes are not always used and it is therefore not always possible to distinguish type-1 and type-2 on the basis of the ICPC coding.

Diagnostic tests

The Diagnostic test table is used to record diagnostic test outcomes. It is designed to facilitate the exchange information between laboratories and different EHR systems . An online code viewer is available at <http://www.consultwijzer.nl/lcw>.

This table is implemented in the EHR system, so the doctor can use it within his system. When a laboratory reports the results of a test, the corresponding

"Labcode" is enclosed. The receiving system also knows this code and places the received result at the appropriate test.

Claims codes

The reimbursement system in the Netherlands is a mix of capitation and fee for service elements. The reimbursable services involved are recorded using 'CTG codes'. CTG-code 13029 for example, indicates that a patient participates in a disease management programme for diabetes.

Prescriptions

The Anatomical Therapeutic Chemical (ATC) Classification System (39) is used for the classification of drugs. It is controlled by the WHO Collaborating Centre for Drug Statistics Methodology (WHOCC). This pharmaceutical coding system divides drugs into different groups according to the organ or system on which they act and/or their therapeutic and chemical characteristics. Each bottom-level ATC code stands for a pharmaceutically used substance, or a combination of substances, in a single indication (or use). A searchable version of the ATC coding system is provided at http://www.whooc.no/atc_ddd_index/.

Appendix B: GPRD database: coding and classifications

Clinical Events

Read terminology was designed by the GP, Dr James Read, for use in his own clinical practice. Following widespread professional endorsement of this terminology, the NHS purchased it in 1990 and decided that it would become the standard for use in the NHS including general practice. The Read system of coding was formally introduced in 1994 and replaced the previously used OXMIS medical codes. Read terminology is a structured hierarchy of both medical and non-medical terms. It covers a number of areas including categories for signs and symptoms, diagnoses, investigations, treatment and therapies, drugs and appliances, occupations and administrative processes. These allow the detailed documentation of patient care whether as a single episode of care or a full electronic patient record.

Read codes therefore offer a comprehensive list of clinical terms that can be used to describe the care and treatment of patients. They were introduced to provide health care professionals with easier access to patient information for the purposes of reporting, research and decision support.

Read codes are Crown Copyright and are distributed by the NHS Terminology Service, which is part of NHS Connecting for Health, on a regular basis. Data in the GPRS is coded using modified version 2 of the Read coding frame. A complete copy of the Read hierarchy is available upon request from the GPRD Team.

Prescriptions

GPs select and record prescription items using the Multilex Product Dictionary (Multilex). Prescription items may be medicinal products, dietary supplements, specialist foods, surgical dressings, or devices and appliances (e.g. space inhalers, syringes). Multilex names and codes, describing the items selected from the Multilex, are associated with therapy events in GPRD.

Under the NHS, it is deemed good practice to prescribe generically. The practice management system interacts with the Multilex to facilitate this process. Each brand name within the Multilex has a generic equivalent, even where there is only one proprietary product available in the UK for a particular formulation of a drug substance. In this latter situation, the generic product listed in Multilex is known as a “theoretical generic product”.

It is therefore critical that all research into drug exposure is initiated at the drug substance level, rather than from a product name level.

(Downloaded from GPRD.com)

Appendix C: TRANSFoRm use case GORD

Introduction

The TRANSFoRm use case 2B: Quality of care and evidence for treatment of gastroesophageal reflux disease (GORD)

Gastroesophageal reflux disease (GERD/GORD) includes a spectrum of disorders mainly caused by the retrograde flow of acid gastric contents from the stomach into the oesophagus, causing symptoms and/or oesophageal mucosal damage. Heartburn and acid regurgitation (a bitter burning taste at the back of the mouth) are the most typical symptoms of GORD. The prevalence of GORD is estimated to 10 % to 20 % in the general population. GORD is treated with proton pump inhibitors (PPI), H₂-blockers or antacid but about 20 % will have persistent or intermittent troublesome symptoms in spite of PPI treatment. GORD is a chronic disease with negative effect on quality of life (QoL), and the disorder is defined as a disease when QoL is affected by the burden of the GORD symptoms. Both for assessment of the quality of clinical care and for research outcomes for GORD it is necessary to be able to measure both symptom burden and QoL easily in symptomatic persons. This is not routinely done in primary health care today.

Purpose and population

We now will elaborate this use case implementing the axioms of the data model that is explained in section 2.

A1: “Data Quality can only be determined within the context of a specified *purpose* and related *knowledge*, and that purpose must be explicitly stated”.

The TRANSFoRm GORD study (use case 2B) is designed as a randomized controlled trial to be performed at several primary health practices across Europe. The research question is “what gives most symptom relief and improvement in quality of life in patients with GORD, on demand or continuous use of proton pump inhibitors?”

In full scale, the study will include 10 countries with 10 primary care practices each, including 10 patients per practice. The patients will be followed for one year.

The second axiom states: “**A2:** A well-specified purpose must determine a *population* of interest.”

The purpose of this use case indicates that we first need to identify patients with GORD. We used the Deliverable 1.1 and searched for GORD codes. Since only ICD-10 diagnoses were given, we had to map those codes to ICPC (NPCD) and Read Codes (GPRD).

In NPCD non-temporary patients, aged 18 or over, with GORD are selected with data concerning diagnosis codes. The ICPC codes to select these patients are: D84 (Disease of the Oesophagus) and D03 (Heartburn/acid regurgitations)

For GPRD the clinical, referral and test records were searched for a code in the GORD code list from D1.1 and the first event date that any such code was mentioned was recorded as the index date. Each patient was assigned a category - which was the minimum category in the above list. So, for example if a patient had a code for category 1 they are classified as category 1, even if they also have a code for categories 2 or 3. Where category is:

1. GORD with a diagnosis (G) code
2. Barrett’s disease
3. No diagnosis (G) code - reflux of heartburn symptoms

Measurement Value Set (MVS)

The measurement value set (MVS) is the set of values along which the quality of data pertaining to one variable is actually measured. The data elements of interest, according to the researchers involved in the TRANSFoRm GORD use case are shown in Table 10.

Entity	CDIM concept	Classification System (LOINC/ICPC)
Demographics		
Age	Human birth instant	
Gender	Biological sex	
Diagnoses		
GORD	Diagnostic_conclusion	ICD-10: K20, K20
Heartburn/Acid regurgitation	Diagnostic_conclusion	ICD-10: R12
Life Style		
Smoking	Clinical history	LOINC:62554-1
Upper gastrointestinal (UGI) endoscopy		
Blood pressure		
BMI		LOINC:39156-5
QoL Questionnaire		

Table 13 Data elements of interest for the GORD Use Case

Metrics

The third axiom states that: “The measurement value set is fully characterized by specifying its completeness, accuracy and correctness. No other characterization is necessary.”

Character	Metric
Diagnose: GORD	
Completeness	% Of patients with code for GORD
Completeness	% Of patients with code for Heartburn or Acid regurgitation
Correctness	% Of patients with code for GORD but more than one prescription for NSAID in the 12 months prior to diagnoses?
Correctness	% Of patients with no code for GORD but with a code for heartburn or acid regurgitation and prescribed PPI at least once each year after diagnosis of heartburn/acid regurgitation
Diagnostic test results: Endoscopy	
Completeness	% Of patients with a code for GORD who have had an upper gastrointestinal (UGI) endoscopy one year prior to index date for the most specific code
Measurements: BMI, QoL, BP,	
Completeness	% Of patients with diagnosis of GORD who have filled in the Quality of Life (QoL) questionnaire
Completeness	% Of patients with a code for GORD with a BMI measurement
Completeness	% Of patients with a code for GORD with a BP measurement
Completeness	% Of patients with a code for GORD with a smoking status measured

Visualising the Quality metrics and Quality Statements

Inter-practice variation in metrics and results of internal correlations are used as measures of data quality.

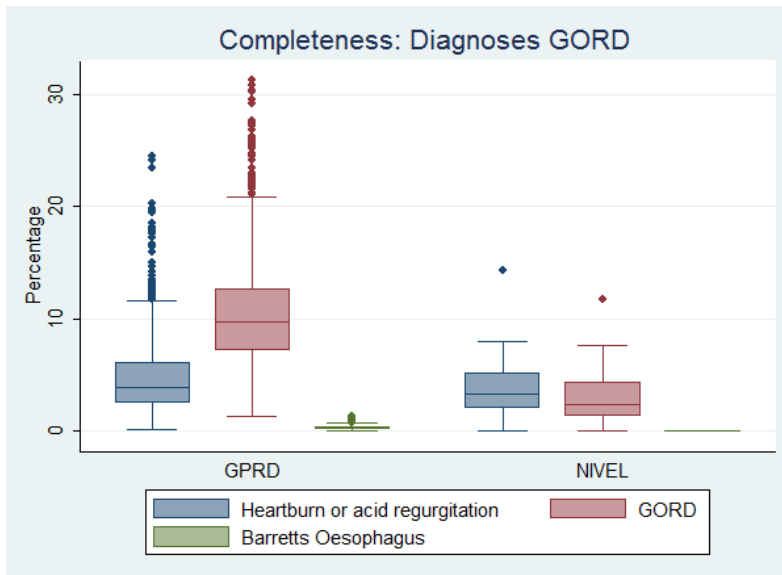


Figure 34 Percentage of non-temporary patients aged 18+ with a code for GORD in 2010 in NPCD and GPRD.

Investigating NSAIDs and PPIs

For this investigation we calculated the proportion of patients with a GORD diagnosis who had a prescription for a) NSAID and b) PPI in the 6 months prior to the first GORD code. We also investigated the proportion that had such a prescription in the 6 months after (and including) the date of diagnosis.

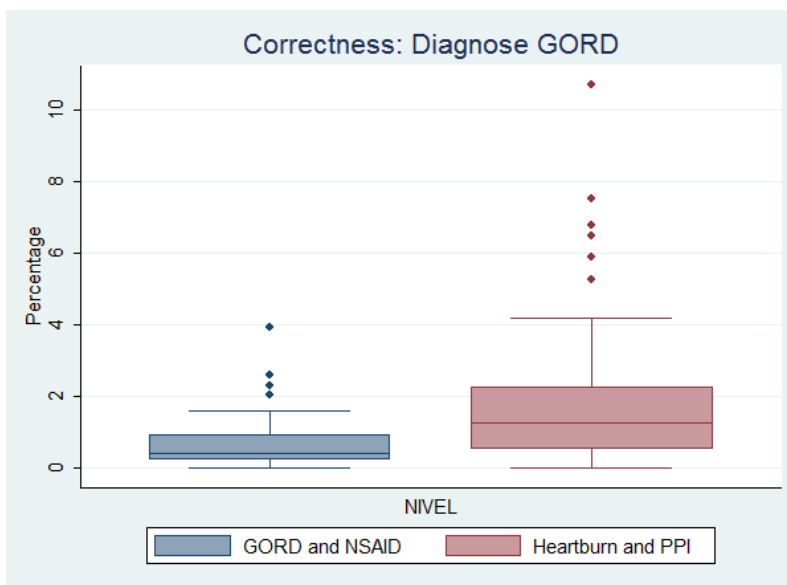


Figure 35 Percentage of GORD patients receiving NSAIDs and percentage of patients with heartburn or acid regurgitation receiving PPI in 2010.

Tests and lifestyle measures for the different categories of coding.

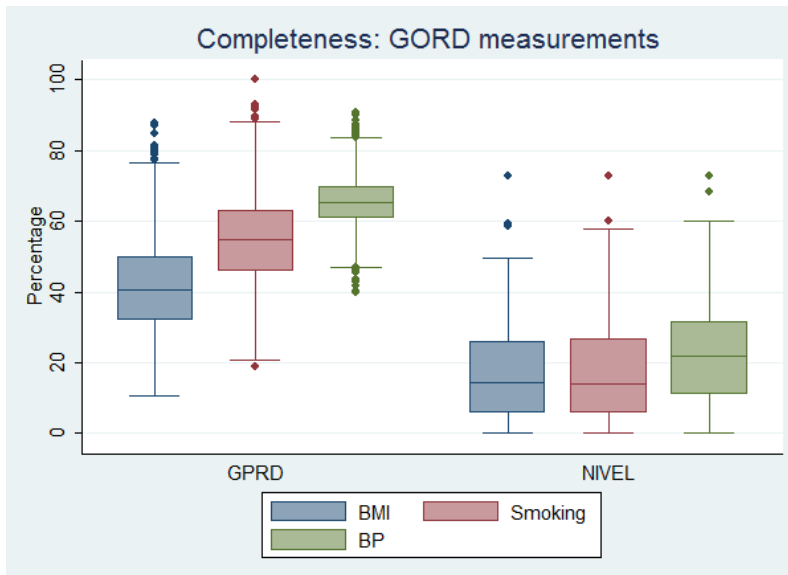


Figure 36 the proportion of patients with GORD within a given practice having at least one measurement of a given type in 2010 calculated for NPCD and GPRD.

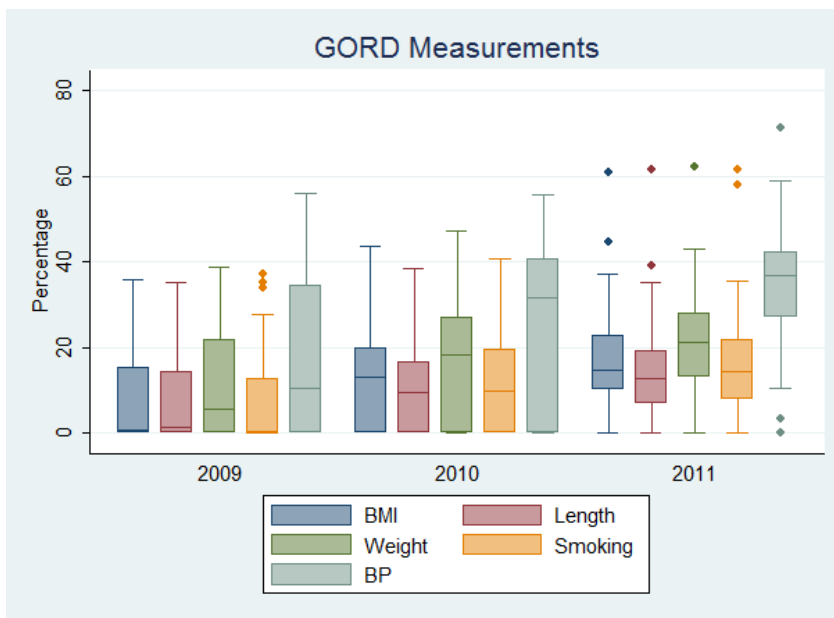


Figure 37 Estimates of completeness for a variety of data entities for the population of patients with GORD in NPCD over the years.

Appendix D: TAM-inspired questionnaire for the Data Quality Tool Evaluation

Questionnaire ID:

The Data Quality Tool (DQT) System Evaluation Questionnaire

Please answer **all** questions in section 1, and then mark **the most appropriate box** for each of the statements in the following sections:

1. Personal Information							
1.1 How old are you? <input type="checkbox"/> 20-25 <input type="checkbox"/> 26-31 <input type="checkbox"/> 32-37 <input type="checkbox"/> 38-43 <input type="checkbox"/> 44-49 <input type="checkbox"/> 50-55 <input type="checkbox"/> 56-61 <input type="checkbox"/> 62+							
1.2 What is your gender? <input type="checkbox"/> Male <input type="checkbox"/> Female							
1.3 What is your specialty? <input type="checkbox"/> Database Designer <input type="checkbox"/> Clinical Researcher <input type="checkbox"/> Data Manager <input type="checkbox"/> Data Controller <input type="checkbox"/> Other, please specify							
1.4 Do you have experience of using Clinical Database Systems? <input type="checkbox"/> Yes <input type="checkbox"/> No							
1.5 If you answered “Yes” to the previous question, how many years?							
1.6 Are you keen on using the DQT System voluntarily? <input type="checkbox"/> Yes <input type="checkbox"/> No							
2. Performance Expectancy	Strongly Disagree	Very Much Disagree	Disagree	Not Sure	Agree	Very Much Agree	Strongly Agree
2.1 I find the DQT System useful in my job, especially when entering and searching data							
2.2 Using the DQT System enables me to complete tasks on a quicker pace							
2.3 Using the DQT System increases my productivity							
2.4 Using the DQT System improves the quality of my work							
2.5 Overall, the DQT System fits well with the way I work and the service I provide							
3. Effort Expectancy	Strongly Disagree	Very Much Disagree	Disagree	Not Sure	Agree	Very Much Agree	Strongly Agree

		Disagree				Agree	
3.1 Interacting with the DQT System has been user-friendly, clear, and understandable							
3.2 Learning to operate the DQT System has been easy for me and did not take a long time							
3.3 I found it easy to become skilful at using the DQT System (for example: applying extensive and/or particular search enquiries)							
3.4 The DQT System facilitates my daily job							
3.5 I believe that interacting with the DQT System does not mentally exhaust me							
4. Social Influence	Strongly Disagree	Very Much Disagree	Disagree	Not Sure	Agree	Very Much Agree	Strongly Agree
4.1 My managers/supervisors encourage me to use the DQT System							
4.2 The long-term benefits of using the DQT System inspire me to use it							
4.3 Within my organization, the management usually communicates their plans of introducing new systems to the staff							
4.4 Within my organization, the management usually offers sufficient training when new systems are introduced							
4.5 Overall, the use of healthcare information technology in my workplace is very important							
5. Facilitating Conditions	Strongly Disagree	Very Much Disagree	Disagree	Not Sure	Agree	Very Much Agree	Strongly Agree
5.1 The training/help option was/is comprehensive enough							
5.2 The use of the DQT System is smooth (no technical errors)							
5.3 The steps followed by the system are logical							

to use, apply, and recall							
5.4 If I got stuck, a sufficient reference or resource is available to help me out							
5.5 The DQT System is easily accessible, responds to queries promptly, and does not disturb my daily routine							
6. DQT-Specific Features	Strongly Disagree	Very Much Disagree	Disagree	Not Sure	Agree	Very Much Agree	Strongly Agree
6.1 DQT system: the number of steps to select practices fit for purpose is reasonable							
6.2 DQT login: logging into the system is quick enough							
6.3 DQT information map: there is enough information about available databases, their country of origin, and owner contact details							
6.4 DQT specific database: for a database of interest, some further general requirements can be selected such as: whether the database is linkable to generic data, to a cancer registry, to drug registry, etc.							
6.5 DQT data quality: the system assesses the completeness of available data in a visual output which is helpful to interpret							
6.6 DQT defining use case: querying data for a certain study using use case, population, and time period is prompt enough							
6.7 DQT specifying use cases: after a use case is defined, further specification can be applied using slide bars on predefined data elements							
6.8 DQT output: the practice ID given for a specified use case is the right output							
7. Technology Anxiety	Strongly Disagree	Very Much Disagree	Disagree	Not Sure	Agree	Very Much Agree	Strongly Agree
7.1 Generally, I like to keep up with new information technology							
7.2 I am open to change my work habits if							

necessary							
7.3 I consider myself technology-oriented due to the nature of my job							
7.4 If I were given enough time and sufficient resources to absorb change, I am likely to accept it							
7.5 Overall, I am not anxious about using information technology							
8. Adaption Timeline	Strongly Disagree	Very Much Disagree	Disagree	Not Sure	Agree	Very Much Agree	Strongly Agree
8.1 I had adequate time to get trained on using the system							
8.2 I received sufficient information and/or training during this time							
8.3 I had enough time to get used the system environment							
8.4 If the system became available, I think I would be able to improve the quality of my work immediately							
8.5 I think I would be confident enough to carry out clinical trials by the help of this system right away if it becomes available							
8.6 Overall, I think I will be allowed enough time to get into the habit of using this system if it becomes available							
9. Behavioral Intention	Strongly Disagree	Very Much Disagree	Disagree	Not Sure	Agree	Very Much Agree	Strongly Agree
9.1 I intend to learn how to use the DQT System to the best of my abilities							
9.2 I intend to use the DQT System as often as needed to perform my work							
9.3 I intend to promote the DQT System to my colleagues if it became available for broader use							
9.4 I intend to suggest ideas to further enhance the DQT System if I had any							

9.5 I prefer to use the DQT System because data can be better preserved, archived, shared, manipulated, and searched electronically.						
10. Your opinion matters. Please leave either a general or a specific comment below if you would like to. If you are concerned about certain sections, just mention their number as a reference point.						
<p>Thank you!</p>						

References

Reference List

- (1) van der Lei J. Use and abuse of computer-stored medical records. *Methods Inf Med* 1991 Apr;30(2):79-80.
- (2) expert.doc. <http://www.expertdoc.nl/nhgdoc/voor-wie>. <http://www.expertdoc.nl/nhgdoc/voor-wie> 2013
- (3) Chan KS, Fowles JB, Weiner JP. Review: electronic health records and the reliability and validity of quality measures: a review of the literature. *Med Care Res Rev* 2010 Oct;67(5):503-27.
- (4) Stirbu-Wagner I, Spies-Dorgelo M, Verheij R. Huisartsenzorg in cijfers: kwaliteitsindicatoren vanuit het HIS? 54 ed. 2011.
- (5) Bij Svd, Khan N, Veen Pt, Roodzant E, Visscher S, Verheij R. De kwaliteit van elektronische verslaglegging door huisartsen gemeten: EPD-scan regio Twente, eindrapport. Utrecht: NIVEL; 2013.
- (6) Gillam S, Siriwardena AN. the quality and outcomes framework. Radcliffe; 2011.
- (7) Dorn T, Yzermans CJ, Guijt H, van der Zee J. Disaster-related stress as a prospective risk factor for hypertension in parents of adolescent fire victims. *Am J Epidemiol* 2007 Feb 15;165(4):410-7.
- (8) van Dijk CE, Verheij RA, Te BH, Spreeuwenberg P, Groenewegen PP, de Bakker DH. Changes in the remuneration system for general practitioners: effects on contact type and consultation length. *Eur J Health Econ* 2013 Feb 28.
- (9) Reeves D, Campbell SM, Adams J, Shekelle PG, Kontopantelis E, Roland MO. Combining multiple indicators of clinical quality: an evaluation of different analytic approaches. *Med Care* 2007 Jun;45(6):489-96.
- (10) Maas J, Verheij RA, de VS, Spreeuwenberg P, Schellevis FG, Groenewegen PP. Morbidity is related to a green living environment. *J Epidemiol Community Health* 2009 Dec;63(12):967-73.
- (11) Fassaert T, Nielen M, Verheij R *et al*. Quality of care for anxiety and depression in different ethnic groups by family practitioners in urban areas in the Netherlands. *Gen Hosp Psychiatry* 2010 Jul;32(4):368-76.
- (12) Florentinus SR, Nielsen MW, van DL, Leufkens HG, Hansen EH, Heerdink ER. Patient characteristics associated with prescribing of a newly introduced drug: the case of rofecoxib. *Eur J Clin Pharmacol* 2005 Apr;61(2):157-9.
- (13) RIVM. Public health surveys and forecasts. 2013.

- (14) Prokosch HU, Ganslandt T. Perspectives for medical informatics. Reusing the electronic medical record for clinical research. *Methods Inf Med* 2009;48(1):38-44.
- (15) Kopcke F, Trinczek B, Majeed RW *et al.* Evaluation of data completeness in the electronic health record for the purpose of patient recruitment into clinical trials: a retrospective analysis of element presence. *BMC Med Inform Decis Mak* 2013;13:37.
- (16) Steinbusch PJ, Oostenbrink JB, Zuurbier JJ, Schaepkens FJ. The risk of upcoding in casemix systems: a comparative study. *Health Policy* 2007 May;81(2-3):289-99.
- (17) NHS audit commission. Information and data quality in th the NHS. 2004.
- (18) Nederlands Huisartsengenootschap. Richtlijn adequate dossiervorming met het elektronisch patientendossier ADEPD. 2009.
- (19) Verdonck P, Strobbe J, Steenackers J *et al.* Aanbevelingen voor goede medische praktijkvoering Het elektronisch medisch dossier. 33(2) ed. 2004.
- (20) Wright A, McCoy AB, Henkin S, Kale A, Sittig DF. Use of a support vector machine for categorizing free-text notes: assessment of accuracy across two institutions. *J Am Med Inform Assoc* 2013 Mar 30.
- (21) Ethier JF, Dameron O, Curcin V *et al.* A unified structural/terminological interoperability framework based on LexEVS: application to TRANSFoRm. *J Am Med Inform Assoc* 2013 Apr 9.
- (22) Kahn MG, Raebel MA, Glanz JM, Riedlinger K, Steiner JF. A pragmatic framework for single-site and multisite data quality assessment in electronic health record-based clinical research. *Med Care* 2012 Jul;50 Suppl:S21-S29.
- (23) Juran JM. *Quality Control Handbook*. New York, NY: McGraw-Hill, 1951; 1951.
- (24) CIHI. the CIHI data quality framework. Ottawa: Canadian Institute for Health Information; 2009.
- (25) Nederlands Huisartsengenootschap. Richtlijn adequate dossiervorming met het elektronisch patientendossier ADEPD. 2009.
- (26) Verdonck P, Strobbe J, Steenackers J *et al.* Aanbevelingen voor goede medische praktijkvoering Het elektronisch medisch dossier. 33(2) ed. 2004.
- (27) Wang RY, Strong DM. Beyond accuracy: what data quality means to data consumers. *Journal of management information systems* 1996;12(4):5-34.
- (28) Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. *J Am Med Inform Assoc* 2013 Jan 1;20(1):144-51.
- (29) Thiru K, Hassey A, Sullivan F. Systematic review of scope and quality of electronic patient record data in primary care. *BMJ* 2003 May 17;326(7398):1070.

- (30) van den Dungen C, Hoeymans N, Schellevis FG, van Oers HJ. Quality aspects of Dutch general practice-based data: a conceptual approach. *Fam Pract* 2013 Jan 10.
- (31) Juran JM. *Quality Control Handbook*. New York, NY: McGraw-Hill, 1951; 1951.
- (32) Hassan SN, Sadek AR, Tahir A, Khunti K, Desombre T, de LS. Evaluating tools to support a new practical classification of diabetes: excellent control may represent misdiagnosis and omission from disease registers is associated with worse control. *Int J Clin Pract* 2012 Sep;66(9):874-82.
- (33) Tate AR, Beloff N, Padmanabhan S, Van Staa T, Williams T. A protocol for characterising data quality in primary care research databases. 2013.
- (34) Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. *J Am Med Inform Assoc* 2013 Jan 1;20(1):144-51.
- (35) Kuchinke W, Van Veen E-B, Delaney BC, Verheij RA, Taweel A, Ohmann C. The discussion on Friday should surround the final model diagram (fig 6) and the text Brendan refers to at the bottom of page 28 [AR5, AA6, AR7]. 2011.
- (36) Verheij R, Zee Jvd. Collecting information in general practice: 'just by pressing a single button'? In: westert GP, Jabaaij L, Schellevis FG, editors. *Morbidity, performance and quality in primary care: Dutch general practice on stage*. Oxon: Radcliffe Publishing, 2006.
- (37) Lamberts H, Wood M, Hofmans-Okkes IM. International primary care classifications: the effect of fifteen years of evolution. *Fam Pract* 1992 Sep;9(3):330-9.
- (38) Soler JK, Okkes I, Wood M, Lamberts H. The coming of age of ICPC: celebrating the 21st birthday of the International Classification of Primary Care. *Fam Pract* 2008 Aug;25(4):312-7.
- (39) WHO collaborating center for Drug Statistics Methodology. *ATC structure_and_principles*. 2013.
- (40) Venkatesh V, Morris MG, Gordon BD, Davis FD. User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly*,2003; 27, 425-478.
- (41) Davis FD.. A technology acceptance model for empirically testing new end-user information systems : theory and results. Massachusetts Institute of Technology, 1985.
- (42) Davis, FD. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*, 1989; 13, 319-340.
- (43) Holden RJ, Karsh B-T. The Technology Acceptance Model: Its past and its future in health care. *Journal of Biomedical Informatics*, 2010; 43, 159-172.

- (44) Algharibi AJ, Arvanitis TN. Adapting the unified theory of acceptance and use of technology (UTAUT) as a tool for validating user needs on the implementation of e-trial software systems. Proceedings of the 25th BCS Conference on Human-Computer Interaction. Newcastle-upon-Tyne, United Kingdom: British Computer Society, 2011.
- (45) Ajzen, The Theory of Planned Behavior. Organisational Behavior and Human Decision Process, 1991; 50, 179-211.
- (46) Thompson RL, Higgins CA, Howell JM. Personal Computing: Toward a Conceptual Model of Utilization. MIS Quarterly, 1991; 15, 125-143.

List of Abbreviations

BMI	Body Mass Index
BP	Blood pressure
CDIM	Clinical Data Integration Model (TRANSFoRm WT6.3)
EHR	Electronic Health Record
GORD	Gastro Oesophagel Reflux Disease
GP	General practitioner
GPRD	General Practice Research Database
H2A	H₂-receptor antagonists
HbA1c	Glycated hemoglobin, hemoglobin A1c
ICD-10	International Classification of Diseases
ICPC	International Classification of Primary Care
LOINC	Logical Observation Identifiers Names and Codes
MVS	Measurement Value Set
NIVEL	Netherland Institute for Health Services Research
NPCD	Netherlands Primary Care Database
NSAID	Non sterpod anti-inflammatory drugs
PPI	Proton pump inhibitors
QoF	Quality and Outcomes framework
QoL	Quality of Life
SNOMED CT	<i>Systematized Nomenclature of Medicine Clinical Terms</i>

