# D5.1 Data Quality Tool

<table>
<thead>
<tr>
<th>Work Package Number:</th>
<th>WP5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Work Package Title:</td>
<td>INTERFACES: USER AND SOFTWARE</td>
</tr>
<tr>
<td>SERVICES (SO5)</td>
<td></td>
</tr>
<tr>
<td>Nature of Deliverable:</td>
<td>Report</td>
</tr>
<tr>
<td>Dissemination Level:</td>
<td>Confidential</td>
</tr>
<tr>
<td>Version:</td>
<td>1</td>
</tr>
<tr>
<td>Delivery Date From Annex 1:</td>
<td>M36</td>
</tr>
<tr>
<td>Principal Authors:</td>
<td>N.A. Khan, M.McGilchrist, S.Padmanabhan, T. Van Staa, R.A. Verheij</td>
</tr>
<tr>
<td>Contributing Authors:</td>
<td>-</td>
</tr>
<tr>
<td>Partner Institutions:</td>
<td>NIVEL, University of Dundee, CPRD</td>
</tr>
</tbody>
</table>

---


Contents

List of Figures ....................................................................................................................... 3
List of Tables .......................................................................................................................... 5
Executive summary ................................................................................................................ 6
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 ....................... 11
  2.3 Conclusion ............................................................................................................... 17
3 Data quality framework ............................................................................................... 18
  3.1 Introduction .............................................................................................................. 18
  3.2 The role of scientific or managerial context .................................................... 19
  3.3 Axiomatic basis – purpose and population ..................................................... 20
  3.4 Population determination .................................................................................. 22
  3.5 Measurement Value Set (MVS) ........................................................................ 22
  3.6 Characterisation of a measurement value set .................................................. 24
  3.7 Metrics .................................................................................................................... 26
  3.8 Visualising the metrics ....................................................................................... 30
  3.9 Contextual provenance of data ....................................................................... 33
  3.10 Data quality statements .................................................................................... 35
  3.11 Tasks for determining data quality ................................................................. 37
4 Determining Data Quality for TRANSFoRm Diabetes Use Case .... 39
  4.1 Databases .............................................................................................................. 39
  4.2 Purpose and population ..................................................................................... 40
  4.3 Measurement Value Set (MVS) ........................................................................ 42
  4.4 Metrics .................................................................................................................... 44
  4.5 Visualising the Quality metrics and Quality Statements ......................... 45
  4.6 Discussion .............................................................................................................. 51
5 TRANSFoRm Data Quality Webtool ................................................................. 53
  5.1 Requirements for the Data Quality Webtool ............................................... 53
  5.2 Architecture of the Data Quality Webtool ...................................................... 54
  5.3 User Interface Design and Workflow of the Data Quality Webtool .... 58
  5.4 Evaluation of the Data Quality Webtool ......................................................... 63
Appendix A: NIVEL primary care database: classifications and coding 65
Appendix B: GPRD database: coding and classifications ................. 67
Appendix C: TRANSFoRm use case GORD ......................................................... 69
References ......................................................................................................................... 76
List of Abbreviations ........................................................................................................ 81

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. ................................................................................................. 12

Figure 2 Illustration of the relationship between the core concepts of the data quality model. A1 – axiom 1, A2 – axiom 2......................................................... 21

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 ......................................................... 31

Figure 4 Box plots visualising the distribution of completeness for variables relevant for diabetes use case................................................................. 32

Figure 5. The data quality model .................................................................................. 37

Figure 6 Percentage of non-temporary patients aged 18+ with diabetes code (ever in their record) in 2009, 2010, and 2011 in the NPCD......................... 46

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......................................................... 46

Figure 8 The proportion of diabetic Retinopathy in the know diabetic population in 2009, 2010 and 2011 (NPCD)......................................................... 47

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................................................................. 48

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. ........... 49

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........................................................................ 50

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......................................................................................... 50
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……………………………………………………………………………………………………... 51
Figure 14 Conceptual Architecture ........................................................................... 54
Figure 15 Quality metrics table ................................................................................. 55
Figure 16 Workflow for selecting practices ............................................................... 57
Figure 17 Authorisation page ..................................................................................... 58
Figure 18 Background information about the available databases ......................... 58
Figure 19 Selection of general requirements for the databases ............................... 59
Figure 20 Selection of database ............................................................................... 59
Figure 21 General information about the selected database ................................... 60
Figure 22 Selection of the use case ....................................................................... 60
Figure 23 Selection of the year .............................................................................. 60
Figure 24 Data elements for the use case .............................................................. 61
Figure 25 Distribution of HbA1c metrics over practices in NPCD in 2010 .......... 61
Figure 26 Result of the query ............................................................................... 62
Figure 27 Possibility to save the query or send a request to the database … 62
Figure 28 Feedback about the data quality metrics ............................................... 63
Figure 29 Percentage of non-temporary patients aged 18+ with a code for GORD in 2010 in NPCD and GPRD................................................................. 73
Figure 30 Percentage of GORD patients receiving NSAIDS and percentage of patients with heartburn or acid regurgitation receiving PPI in 2010....... 74
Figure 31 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........................................................................................................ 74
Figure 32 Estimates of completeness for a variety of data entities for the population of patients with GORD in NPCD over the years……………….. 75
List of Tables

Table 1 Definitions for completeness, accuracy and correctness and consistency. ............................................................... 25
Table 2 Character as defined in this report and possible equivalent terms in the literature ....................................................... 26
Table 3 The two basic operations proposed for the data quality model: counting and enumeration. ................................................................. 28
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. .......................... 28
Table 5 Categories of factors that influence the generation and transmission of data in different zones of the TRANSFoRm zone model (35). .......... 34
Table 6 The application of thresholds to each metric provides a specification of data quality for a study ................................................................. 36
Table 7 Basic characteristics of NIVEL Primary Care Database and General Practice Research database ............................................................. 40
Table 8 Data elements of interest for the Diabetes Use Case .............. 44
Table 9 Quality Metrics for Diabetes Use Case ............................................................. 45
Table 10 Data elements of interest for the GORD Use Case ............... 71
Table 11 Quality Metrics for GORD Use Case ............................................................. 72
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.
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 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

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
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 user who finds this event in a research file and starts using it in a query. This whole process is depicted in Figure 1.
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.
  o 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.
  o 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 may 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 it the fact that data isn’t the
same as information and that there is analyses 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

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 (22).

Two statements are frequently cited regarding data quality: “data should be used only for the purpose for which they were collectes” (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 must be rejected if we are to entertain the very idea of re-use of routine data. The second offers only a starting point for a framework; it 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 included both human judgement and computable operations that together permit the assessment of the quality of a given dataset in the data quality web tool.

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 a problem as part of the background knowledge is derived from the data under study (eg. GPRD and NIVEL data are used to generate public health information).

For example, for the population of patient 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 kind of 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 to be 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 reflects 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 paramount. This may result in what might 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:

<table>
<thead>
<tr>
<th>Req/Resp</th>
<th>CDIM concept</th>
<th>Example value(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Request</td>
<td>Laboratory_test_ordering</td>
<td>LOINC: 59261-8</td>
</tr>
<tr>
<td>Request</td>
<td>Measurement unit label</td>
<td>UCUM: mmol/mol</td>
</tr>
<tr>
<td>Response</td>
<td>Scalar measurement datum (laboratory finding)</td>
<td>53</td>
</tr>
</tbody>
</table>

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:

<table>
<thead>
<tr>
<th>Req/Resp</th>
<th>CDIM concept</th>
<th>Example value(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Response</td>
<td>Diagnostic_conclusion</td>
<td>ICD10: E11</td>
</tr>
</tbody>
</table>

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 Table 1 where we map these three characters to the dimensions found in the literature) or relate to the operations described in section 3.7.
<table>
<thead>
<tr>
<th>Character</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Completeness</td>
<td>On the completion of a measurement or judgemental process, the <em>degree</em> 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.</td>
</tr>
<tr>
<td>Accuracy</td>
<td>The <em>degree</em> 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).</td>
</tr>
<tr>
<td>Correctness</td>
<td>The <em>degree</em> to which the recorded value conforms to the scientific and managerial context.</td>
</tr>
<tr>
<td></td>
<td>For example, does a diagnosis code conform to the context?</td>
</tr>
<tr>
<td>Consistency</td>
<td>The <em>degree</em> to which each of the characters of correctness, accuracy and correctness are unchanged over time, where time refers to the moment at which these characters are assessed.</td>
</tr>
</tbody>
</table>

*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 to mean complete and correct.

<table>
<thead>
<tr>
<th>Character</th>
<th>Dimension(s) mentioned in the literature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Completeness</td>
<td>Completeness, availability, missingness, omission, presence, sensitivity</td>
</tr>
<tr>
<td>Accuracy</td>
<td>Granularity, level of detail</td>
</tr>
<tr>
<td>Correctness</td>
<td>Error, validity, plausibility, believability, trustworthiness, positive predictive value, agreement, concordance.</td>
</tr>
<tr>
<td>Consistency</td>
<td>Reliability</td>
</tr>
</tbody>
</table>

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 *measurable* determinant is time.

The counting and enumeration operations we propose are listed below:

<table>
<thead>
<tr>
<th>Operation</th>
<th>Applies to character</th>
<th>Example metric</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Count</strong></td>
<td>Completeness</td>
<td>HbA1c values are counted and the counts expressed as ratios to the population denominators and profiled by a quality determinant such as practice or period of time (context). See Figure 3.</td>
</tr>
<tr>
<td><strong>Enumerate (distribution)</strong></td>
<td>Completeness</td>
<td>Value distributions are inspected for missing values 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.</td>
</tr>
<tr>
<td>Correctness</td>
<td>Diagnosis codes</td>
<td>Diagnosis codes are enumerated and <em>cross-checked</em> against scientific context such as ICD10 diabetes codes. Counts are provided for those within the required classifications and those falling outside the classifications. Age values might be compared against a</td>
</tr>
</tbody>
</table>
known profile with known bounds for a particular population.

Accuracy

The enumerated values or codes are cross-checked against scientific and managerial context for the degree of resolution available.

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 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).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Practice code (quality determinant)</th>
<th>Diagnosis code</th>
<th>Count (% practice population)</th>
<th>Context</th>
<th>Node designation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>14</td>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>E10-E14</td>
<td>194 (1%)</td>
<td>ICD10</td>
<td>Branch (level n)</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>E10</td>
<td>1837 (9%)</td>
<td>ICD10</td>
<td>Branch (n+1)</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>E11</td>
<td>19041 (90%)</td>
<td>ICD10</td>
<td>Branch (n+1)</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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 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).

<table>
<thead>
<tr>
<th>Factor category</th>
<th>Care zone (data recording)</th>
<th>Non-care (Linkage) zone (ETL: EHRs to repository)</th>
<th>Research zone (Data counts from platform)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Healthcare system</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Professional Guidelines</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Knowledge management</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Technical processes</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Quality control</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Governance</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Patient/provider behavior</td>
<td>X</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

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.

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.
<table>
<thead>
<tr>
<th>Databases</th>
<th>NIVEL Primary care database</th>
<th>General Practice Research Database</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Size of database</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Country</td>
<td>Netherlands</td>
<td>England</td>
</tr>
<tr>
<td><strong>Number of practices</strong></td>
<td>300 GP practices</td>
<td>650 GP practices</td>
</tr>
<tr>
<td><strong>Number of patients</strong></td>
<td>1 million</td>
<td>12 million</td>
</tr>
<tr>
<td><strong>Number of EHR brands</strong></td>
<td>7 (Medicom, Omnihis, Promedico ASP, Promedico VDF, MicroHIS, Mira)</td>
<td>1 (Vision)</td>
</tr>
<tr>
<td><strong>Frequency of data upload</strong></td>
<td>Twice a year</td>
<td>Every month</td>
</tr>
<tr>
<td><strong>Content</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diagnoses/symptoms</td>
<td>ICPC-1</td>
<td>Read codes v2</td>
</tr>
<tr>
<td>Diagnostic tests, lab results</td>
<td>NHG codes</td>
<td>Read codes v2</td>
</tr>
<tr>
<td>Prescriptions</td>
<td>ATC</td>
<td>Multilex</td>
</tr>
<tr>
<td>Claims</td>
<td>NZa codes</td>
<td></td>
</tr>
<tr>
<td>Procedures</td>
<td>NZa codes</td>
<td></td>
</tr>
<tr>
<td>Data available from</td>
<td>1995-2012</td>
<td>1987-2012</td>
</tr>
<tr>
<td><strong>Health system characteristics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Health service characteristics</td>
<td>Managed competition</td>
<td>National Health Service</td>
</tr>
<tr>
<td>Basis for reimbursement</td>
<td>Mix of capitation and fee for service</td>
<td>Capitation and quality of care</td>
</tr>
<tr>
<td>List system</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Other</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Record linkage possibilities</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

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
<table>
<thead>
<tr>
<th>Entity</th>
<th>CDIM concept</th>
<th>Classification System (LOINC/ICPC)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Demographics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Age/Year of Birth</strong></td>
<td>Human birth instant</td>
<td></td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td>Biological sex</td>
<td></td>
</tr>
<tr>
<td><strong>Race/Ethnicity</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Diagnoses</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Time of Diagnoses</strong></td>
<td>Diagnostic conclusion instant</td>
<td>ICPC codes: T90 (T90.2?) SNOMED-CT: 44054006 ICD-10: non-insulin dependent diabetes mellitus (obsolete): E11</td>
</tr>
<tr>
<td><strong>Diagnose: Type-2 Diabetes</strong></td>
<td>Diagnostic_ conclusion</td>
<td>ICPC: F83.01 ICD-10: E11.3 (E11.3A, E11.3B, E11.3C differentiates retinopathy from other eye complications)</td>
</tr>
<tr>
<td><strong>Diabetic Retinopathy</strong></td>
<td>Diagnostic_ conclusion</td>
<td>ICPC: U99.1 ICD10: E11.2 E11.2A and E11.2B differentiates nephropathy from other kidney complications</td>
</tr>
<tr>
<td><strong>Diabetes Nephropathy</strong></td>
<td>Diagnostic_ conclusion</td>
<td>ICPC: U99.1 ICD10: E11.2 E11.2A and E11.2B differentiates nephropathy from other kidney complications</td>
</tr>
<tr>
<td><strong>Family history of diabetes</strong></td>
<td>Diagnostic_ conclusion</td>
<td>ICPC: U99.1 ICD10: E11.2 E11.2A and E11.2B differentiates nephropathy from other kidney complications</td>
</tr>
<tr>
<td><strong>Life Style</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Physical activity habits/sedentarism</strong></td>
<td>Clinical history</td>
<td></td>
</tr>
<tr>
<td><strong>Smoking status</strong></td>
<td>Clinical history</td>
<td>LOINC: 62554-1</td>
</tr>
<tr>
<td><strong>Diagnostic test results</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Hba1c (values and dates)</strong></td>
<td>Scalar measurement datum laboratory finding</td>
<td>LOINC: 59261-8 LOINC: 4548-4</td>
</tr>
<tr>
<td><strong>Blood glucose (values and dates)(fasting)</strong></td>
<td>Scalar measurement datum laboratory finding</td>
<td>LOINC: 1556-0 LOINC: 14771-0 LOINC: 1558-6</td>
</tr>
<tr>
<td><strong>Serum creatinine</strong></td>
<td>scalar measurement datum laboratory finding</td>
<td>LOINC: 2160-0 LOINC: 14682-9</td>
</tr>
</tbody>
</table>
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 researcher 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 web tool.

<table>
<thead>
<tr>
<th>Character</th>
<th>Metric</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Diagnoses: Type-2 Diabetes</strong></td>
<td></td>
</tr>
<tr>
<td>Correctness</td>
<td>% Of patients with a Metformin prescription and a diabetic code recorded</td>
</tr>
<tr>
<td>Accuracy</td>
<td>% Of patients with a diabetes type-2 code of all the known patients with diabetes in the practice population (denominator is known diabetes patients)</td>
</tr>
<tr>
<td>Consistency</td>
<td>Above metrics are profiled over time.</td>
</tr>
</tbody>
</table>

| 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/m2, 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.
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**

![Box plot showing completeness of diagnoses for Type-2 Diabetes](image)

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

![Graphs showing percentage of registered patients](image)

**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
<table>
<thead>
<tr>
<th>Data entity</th>
<th>Metric</th>
<th>Quality Statements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnose</td>
<td>Completeness through distribution</td>
<td>Include practices for which &gt;5% of the practice population have a diabetes code recorded</td>
</tr>
<tr>
<td>Diagnose</td>
<td>Correctness through cross-checking</td>
<td>Include practices for which &gt;95% of patients with a metformin prescription also have a diabetes diagnoses recorded</td>
</tr>
<tr>
<td>Diagnose</td>
<td>Accuracy</td>
<td>Include practices for which &gt;90% of the patients have a type two diagnoses code</td>
</tr>
</tbody>
</table>

**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](image)

*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 even though 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.

<table>
<thead>
<tr>
<th>Data entity</th>
<th>Metric</th>
<th>Quality Statements</th>
</tr>
</thead>
<tbody>
<tr>
<td>HbA1c</td>
<td>Completeness through counts</td>
<td>Include practices for which &gt;90% of diabetic patients have a value</td>
</tr>
<tr>
<td>HbA1c</td>
<td>Correctness through enumeration</td>
<td>Include practices for which &gt;70% of all HbA1c values are between 0-300 mmol/mol Or 0-100%</td>
</tr>
<tr>
<td>HbA1c</td>
<td>Accuracy through enumeration</td>
<td>Include practices for which &gt;10% of HbA1c values are accurately written in decimal</td>
</tr>
</tbody>
</table>

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.
<table>
<thead>
<tr>
<th>Data entity</th>
<th>Metric</th>
<th>Quality Statements</th>
</tr>
</thead>
<tbody>
<tr>
<td>BMI</td>
<td>Completeness through counts</td>
<td>Include practices for which &gt;80% of diabetic patients have a value</td>
</tr>
<tr>
<td>BMI</td>
<td>Correctness through enumeration</td>
<td>Include practices for which &gt;30% of all BMI values are between 15-50 kg/m²</td>
</tr>
<tr>
<td>BMI</td>
<td>Accuracy through enumeration</td>
<td>Include practices for which &gt;10% of BMI values are accurately written in decimal</td>
</tr>
</tbody>
</table>

*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](image)

*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/](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.

![Conceptual Architecture](image)

**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.

![Quality metrics table](image)

**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 we 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 screenshots 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: n.khan@nivel.nl
Password: 
Login

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.

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.

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.
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.

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

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.
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.

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).
5.4 Evaluation of the Data Quality Webtool

To evaluate the webtool we used a TAM-questionnaire (TAM: technology acceptance model). The TAM-questionnaire was given to the end users, in this case the principal investigators of each TRANSFoRm use case, to measure the acceptance of the data quality webtool. We didn't think it was possible to give it to other researchers out side the project since the metrics that are implemented in the webtool are dependant on the use case. Besides some Personal Information, there were questions about Performance Expectancy, Effort Expectancy, Social Influence, Facilitating Conditions, Technology Anxiety, Adoption Timeline and Behavioral Intention.

Overall the users expected the tool to improve their work performance, increase their productivity and get them better control over their work. They didn’t expect the tool to help them acquire transferable skills.

The researchers had differing views about the interaction with the tool, but were able to identify relevant data easily.

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. The researchers were not anxious about using technology.
Overall the users tend to adapt the tool, if allowed enough time to get into the habit of using this tool if it becomes available. They also intend to do more research studies if the system proved to substantially facilitate so.
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.whocc.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 form 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: “\textbf{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

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

Table 10 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.”

<table>
<thead>
<tr>
<th>Character</th>
<th>Metric</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnose: GORD</td>
<td></td>
</tr>
<tr>
<td>Completeness</td>
<td>% Of patients with code for GORD</td>
</tr>
<tr>
<td>Completeness</td>
<td>% Of patients with code for Heartburn or Acid</td>
</tr>
</tbody>
</table>

D5.1 Data Quality Tool FP7-247787-TRANSFoRm
Table 11 Quality Metrics for GORD Use Case

<table>
<thead>
<tr>
<th>Quality Metrics</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Correctness</strong></td>
<td>% Of patients with code for GORD but more than one prescription for NSAID in the 12 months prior to diagnoses?</td>
</tr>
<tr>
<td><strong>Correctness</strong></td>
<td>% 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</td>
</tr>
<tr>
<td><strong>Diagnostic test results: Endoscopy</strong></td>
<td>% 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</td>
</tr>
<tr>
<td><strong>Measurements: BMI, QoL, BP,</strong></td>
<td>% Of patients with diagnosis of GORD who have filled in the Quality of Life (QoL) questionnaire</td>
</tr>
<tr>
<td><strong>Completeness</strong></td>
<td>% Of patients with a code for GORD with a BMI measurement</td>
</tr>
<tr>
<td><strong>Completeness</strong></td>
<td>% Of patients with a code for GORD with a BP measurement</td>
</tr>
<tr>
<td><strong>Completeness</strong></td>
<td>% Of patients with a code for GORD with a smoking status measured</td>
</tr>
</tbody>
</table>
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 29 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 30 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 31 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 32 Estimates of completeness for a variety of data entities for the population of patients with GORD in NPCD over the years.
References

Reference List


(13) RIVM. Public health surveys and forecasts. 2013.


(24) CIHI. the CIHI data quality framework. Ottawa: Canadian Institute for Health Information; 2009.


(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,


## List of Abbreviations

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