Core Standards of the EUBIROD Project*

Defining a European Diabetes Data Dictionary for Clinical Audit and Healthcare Delivery

S. G. Cunningham[1]; F. Carinci[2,3]; M. Brillante[1]; G. P. Leese[1]; R. R. McAlpine[1]; J. Azzopardi[4]; P. Beck[5]; N. Bratina[6]; V. Bocquet[7]; K. Doggen[8]; P. K. Jarosz-Chobot[9]; M. Jecht[10]; U. Lindblad[11]; T. Moulton[12]; Ž. Metelko[13]; A. Nagy[14]; G. Olympios[15]; S. Pruna[16]; S. Skeie[17]; F. Storms[18]; C. T. Di Iorio[19]; M. Massi Benedetti[2]

[1]University of Dundee, Scotland; [2]Hub for International Health Research, Italy; [3]University of Surrey, United Kingdom; [4]University of Malta, Malta; [5]Joanneum Research, Austria; [6]University Children's Hospital Ljubljana, Slovenia; [7]Centre Hospitalier de Luxembourg, Luxembourg; [8]Scientific Institute of Public Health, Belgium; [9]Medical University of Silesia, Poland; [10]Havelhöhe Hospital, Germany; [11]University of Gothenburg, Sweden; [12]Adelaide and Meath Hospital, Ireland; [13]Vuk Vrhovac University Clinic for Diabetes, Croatia; [14]University of Debrecen, Hungary; [15]Ministry of Health, Cyprus; [16]Telemedica Consulting, Romania; [17]NOKLUS, Norway; [18]Dutch Institute for Healthcare Improvement (CBO), The Netherlands; [19]Serectrix snc, Italy

Keywords

Data dictionary, standardised definitions, common dataset, interoperability, diabetes

Summary

Background: A set of core diabetes indicators were identified in a clinical review of current evidence for the EUBIROD project. In order to allow accurate comparisons of diabetes indicators, a standardised currency for data storage and aggregation was required. We aimed to define a robust European data dictionary with appropriate clinical definitions that can be used to analyse diabetes outcomes and provide the foundation for data collection from existing electronic health records for diabetes.

Methods: Existing clinical datasets used by 15 partner institutions across Europe were collated and common data items analysed for consistency in terms of recording, data definition and units of measurement. Where necessary, data mappings and algorithms were specified in order to allow partners to meet the standard definitions. A series of descriptive elements were created to document metadata for each data item, including recording, consistency, completeness and quality.

Results: While datasets varied in terms of consistency, it was possible to create a common standard that could be used by all. The minimum dataset defined 53 data items that were classified according to their feasibility and validity. Mappings and standardised definitions were used to create an electronic directory for diabetes care, providing the foundation for the EUBIROD data analysis repository, also used to implement the diabetes registry and model of care for Cyprus.

Conclusions: The development of data dictionaries and standards can be used to improve the quality and comparability of health information. A data dictionary has been developed to be compatible with other existing data sources for diabetes, within and beyond Europe.

Correspondence to:
Dr Scott G. Cunningham
Clinical Technology Centre, Level 7
Ninewells Hospital
Dundee, DD1 9SY
UK
E-mail: scott.cunningham@nhs.net

* Supplementary online material published on our website http://dx.doi.org/10.3414/ME15-01-0016

Methods Inf Med 2016; 55: 166–176
http://dx.doi.org/10.3414/ME15-01-0016
received: January 29, 2015
accepted: August 10, 2015
epub ahead of print: December 15, 2015

1. Introduction

Non-communicable diseases (NCDs) are currently addressed as a global priority for public health monitoring [1]. Among these, diabetes represents a condition that is taken very seriously by policy makers, but seems still difficult to report consistently over time, both nationally and internationally [2].

A global diabetes prevalence of 8.3% was reported in 2013 [3]. Accurate reporting is hampered by the size of the target population, the difficult comparability of classification systems [4] and changing case definitions [5]. In 2014, key diabetes indicators (e.g. complications) were still not available for international comparisons. Since the St. Vincent Declaration [6], a series of initiatives aimed at resolving the problem were conducted in Europe.

DIABCARE [7], promoted by the International Diabetes Federation (IDF) and World Health Organization (WHO) offices of Europe was the first attempt to aggregate diabetes data from local databases for international benchmarking. Its "Basic Information Sheet" represented for many years an influential reference for the evolution of indicator monitoring.

In the early 2000s, the European Commission (EC) financed the projects "EUropean Diabetes Indicator Project" (EUDIP) [8] and "EUropean Core Indicators in Diabetes" (EUCID) [9], with the aim of estab-

lishing a system of diabetes indicators across the European Union (EU). The "EU Report on Major and Chronic Diseases" [10] provided initial terms of reference, highlighting the limited comparability of methods and data elements collected at national level. Between 2005–2008, the EU funded "Best Information through Regional Outcomes" (BIRO) [11] carried out an in-depth assessment of needs, opportunities and barriers of diabetes information in seven countries, specifying a common infrastructure through better integration of regional data collection. The project standardised many different aspects, including measures collected in diabetes registries, analytical techniques and presentation of indicators.

All operations were automatically supported by privacy-enhanced open source software, the "BIRO system" [12], using a split procedure that processed sensitive data at the level of each data source, while exchanging micro-aggregates beyond their boundaries [13].

In 2008, the EC provided further support for the project "EUropean Best Information through Regional Outcomes in Diabetes" (EUBIROD, 2008–2012) [14], with the aim of completing and using the BIRO system for the production of diabetes indicators. In EUBIROD, standardised definitions were regarded as a fundamental prerequisite to improve the quality, relevance, consistency, availability and comparability of regional and national information across different jurisdictions.

A data dictionary, or metadata repository, was considered of interest for its value as a "centralized repository of information about data e.g. meaning, relationships to other data, origin, usage, and format" [15]. Specific examples in the health sector adopting the standard ISO/IEC 11179 included the Australian Metadata Online Registry (MeteOR) [16], the US Health Information Knowledgebase [17] and the National Cancer Data Standards Repository [18].

In the context of diabetes, a European Data Dictionary might support a range of connectivity applications of use for research and disease management. This paper aims to present how EUBIROD developed a specific diabetes data dictionary responding to the following questions:

- which data elements are necessary to compute a core set of diabetes indicators in Europe?
- what data sources are available to deliver results on a routine basis?
- to what extent are data elements complete, consistent and standardised in the available data sources?
- how can one define standardised data elements from the existing definitions adopted in diabetes registries?

2. Methods

The data dictionary was the result of a series of procedures carried out in two separate steps.

2.1 Step 1. The BIRO Common Dataset

The work plan of the BIRO project included precise indications on the preparation of the data dictionary. A mix of evidence-based and feasibility criteria were agreed for the selection of diabetes-related parameters to be routinely collected from multiple datasets available in participating centres.

Box 1 Characteristics considered for each parameter

- Parameter code
- Clinical or agreed definition
- Published source of the definition
- Database storage type (text, integer, decimal, date time, etc)
- Value to define the length a text value
- Unit of measurement for the data item (e.g. m, mmol/l, kg/m^2, %)
- Enumerated Types: for items with a finite number of categories (e.g. diabetes type)
- Upper boundary threshold (if relevant)
- Lower boundary threshold (if relevant)
- Whether or not the local recording is mandatory
- Recommended clinical guideline value
- Published source of the guideline value

The BIRO Clinical Review [19] delivered an initial list, based on a systematic assessment of the scientific literature of diabetes indicators, classified according to scientific soundness, reliability and feasibility. A core set of indicators were included in a reference framework, taking into account different domains considered relevant for diabetes monitoring. The list was initially assessed and periodically revised by partners of the BIRO and EUBIROD projects, including endocrinologists, epidemiologists, statisticians and IT experts.

The BIRO Common Dataset [20, 21] translated the results of the Clinical Review, complemented by further investigation and cross-reference against the existing classifications, producing technical specifications for subsequent data collection and analysis.

►Box 1 shows the characteristics considered for each parameter. The International System of Units (SI) [22] was selected as an agreed standard for numerical values. The consistency of evidence-based definitions was compared against current standards adopted by selected partners of the BIRO project:

- DiabCare [7]
- Forum for Quality Systems in Diabetes Care (FQSD) [23]
- The Scottish Diabetes Core Dataset [24]
- Umbria Diabetes Register (PROMODR) [25]

The definitions for each data item were compared at face-value to agree a common and clinically relevant standard in discussion with all partners. Further comparisons were made with definitions endorsed by the Australian Diabetes, Obesity and Lifestyle (AusDiab) [26] study, adding a view beyond European borders.

The BIRO Common Dataset [27] was finally released with full XML specifications [28] in July 2009. The specifications were implemented into the first prototype of the BIRO software, used to test different choices of local definitions against the commonly agreed structure.

All specifications, including the choice of clinical parameters and quality and outcomes indicators, were further refined based on the results of an independent multidisciplinary evaluation, carried out as

Country	Diabetes Register
Austria*	Forum for Quality Systems in Diabetes Care, Styria
Croatia	National Diabetes Register
Cyprus*	Larnaca Region
Germany	Forum for Quality Systems in Diabetes Care, Rheinland-Pfalz
Hungary	GPMSSP Database
Ireland	AMNCH, Tallaght
Italy*	Regione Umbria and the SID National Network
Luxembourg	Pediatric Clinic Database
Malta*	Mater Dei Hospital
Norway*	National Diabetes Register
Poland	Silesia University Hospital
Romania*	Bucharest Diabetes Database
Slovenia	Type 1 Childhood Diabetes Register
Sweden	Skaraborg Primary Care Diabetes Database
United Kingdom*	DARTS, Tayside, Scotland

*Original partners of the BIRO Project

Table 1 EUBIROD Diabetes Registers contributing to the dataset analysis

an integral part of the workplan of both BIRO and EUBIROD.

2.2 Step 2. The EUBIROD Survey

The EUBIROD project aimed to apply the initial scheme in diverse settings, considering methodological issues e.g. the need for proper references for denominators (both in the diabetic and general population); the state transition from healthy to diagnosed, and migrations over time; structural characteristics, including the geographical location of data sources and the fine partitioning of catchment areas; etc.

The Data Dictionary was extended to take stock of different characteristics (data source, population, subjects) available from multiple sources. The consistency of standard definitions with local practices was assessed to ensure proper functioning of analytical software through a pilot survey conducted across EUBIROD diabetes registers (▶Table 1).

The assessment included a brief description of the local data source (source type e.g. disease management programme or regional/national diabetes register; health workforce and catchment area), and a series of criteria to compare parameters against EUBIROD definitions (▶Table 2). The direct collaboration of data custodians allowed refining the structure of meta-data under real life conditions.

An electronic interface (the "Meta-registry of Diabetes Data Sources") was developed to assist data custodians in classifying each item in a "core list" extracted from the data dictionary, according to specified criteria (▶Figure 1). The meta-registry included the following sections: structure (site header and profile datasets), local data items (implementation of common and activity datasets), privacy performance assessment [29].

The meta-registry allowed for a single custodian to manage multiple data sources, each identified by a "BIRO unique data source identifier" (BUDSI) that was assigned by the user as a combination of codes from the official Nomenclature of Territorial Units for Statistics (NUTS) [30]. This way, each data source could be uniquely linked to a specific catchment area, from which the reference population was assumed to be exclusively attributable to a specific location. The meta-registry is available for further use by current and future partners, and is open to the public for a quick evaluation [31].

Recorded	Yes/no
Mandatory	Mandatory/routine
Consistency	Consistency with EUBIROD definition • High: Exact match • Medium: Minor discrepancy – e.g. Source units require mapping • Low: Major discrepancy – e.g. mapping unavailable • Data Item Unavailable
Completeness	Expressed as a percentage • 0–100%
Overall Quality Score	A value judgement on the ability for the data source to provide complete and consistent data in line with the definition • High: Can provide complete and consistent data • Medium: Minor completeness and consistency issues • Low: Major completeness and consistency issues • Data Item Unavailable
Comments	Describes any further information known about the data item at source that may affect longitudinal analysis or data presentation.

Table 2 Criteria used to assess each EUBIROD data item

2.3 Metadata Collection

Metadata were collected through the survey application using two main sections. The initial section on structural characteristics was used to describe general characteristics of each data source. An additional

Figure 1 BIRO data validation questionnaire in the EUBIROD meta-registry of diabetes data sources

section on "local data items" included a sequential questionnaire looping through all EUBIROD data items for comment and stratification for a specific data source. Adherence to several criteria was requested of the designated data manager (▶Box 2).

Responses requested as boolean items (recorded, mandatory), quantitative criteria (consistency, completeness and overall quality score) and qualitative information (comments) were analysed to identify commonalities and inconsistencies between the agreed common standards and the local data items captured at source by each partner.

Qualitative information were used to refine the dictionary in ways that increased comparability across centres e.g. consistent formats or coding. Boolean and quantitative criteria were used to carry out a descriptive statistical analysis taking each data item as the unit of interest.

Parameters calculated for each item, except for "percentage recorded", were computed by taking into account only records with data item recorded (e.g. mean completeness of a data item computed over data sources with item recorded).

The computed quality score was used to extract an overall user-derived cut-off of high validity, calculated as the upper quartile in the distribution of percentage of data sources with a "high overall quality score". The percentage recorded and computed quality score were used as axes in a graphical plot of feasibility vs validity for each data item. The cut-off of high validity and feasibility was used to partition the plane into four quadrants, defined by the combination of high/low coordinates.

The survey was conducted in August 2010 [32] and completed at each partner site by domain experts including data managers, clinicians and technology experts. All results were extracted from the meta-registry database to a csv file that was loaded into a R [33] program, used to deliver the descriptive analyses for the production of the present paper.

The EUBIROD data dictionary was progressively implemented to include additional information on population data, data source structures, and indicators considered for the first EUBIROD diabetes report in March 2012. Further refinements and consistency checks were conducted up to the present date.

3. Results

A total of 53 EUBIROD data items were included in the survey of diabetes-related local data items. For most data items, the overall quality score was approximately close, albeit slightly lower, to the average of completeness and consistency. For this reason, we regarded the computed quality score as a more objective representation of overall validity. The subjective value expressed by data custodians was used to identify a reasonable cut-off for the bivariate partition into quadrants. The upper quartile of the overall quality score assigned by data custodians was equal to 73%. A rounded empiric value of 75% was considered as a reasonable cut-off for both validity and feasibility.

Results of feasibility vs validity are shown in ▶Figure 2. The display can be used to classify diabetes-related data items according to quadrants indicated in the figure. Results for individual items, ordered by quadrant, percentage recorded and computed quality score, are shown in ▶Appendix 1.

Box 2 Parameters computed from questionnaire results for each data item

- Percentage of data sources with data item recorded ("percentage recorded")
- Percentage of data sources with data item mandatory ("percentage mandatory")
- Percentage of data sources with consistency of data item equal to "High"
- Mean completeness
- Percentage of data sources with overall quality score equal to "High" ("high overall quality score")
- Average of mean completeness and consistency observed in the sample ("computed quality score")

On average, over two thirds (69.4%) of the items identified in the initial clinical review were found to be already recorded in EUBIROD data sources. The average percentage at high feasibility (quadrants IV, II) was equal to 88.1% vs 49.9% in the others. Items were considered mandatory in over three quarters (76.7%) of all data sources. However, the best quadrant included items that were considered as mandatory less frequently (62.4% for quadrant IV vs 79.1%, 83.9% and 84.5% for quadrants III-II-I respectively).

Different results were observed for the individual components of validity: higher levels of completeness were more frequent than consistency (79.2% vs 72.2%), while the variability across quadrants ranged between 62.6% –72.2% for consistency, as opposed to 72.4% –79.2% for completeness. These results seem to indicate that consistency as a criterion might better discriminate validity, and can be preferred over completeness to guide choices in the dictionary.

The results were used to revise the structure of the EUBIROD framework for diabetes indicators and to deliver the data standards adopted for the EUBIROD Data Dictionary.

The final version of the EUBIROD framework is presented in ▶Table 3. It involves five "main" sections (demographic and clinical characteristics, health system, population and risk adjusted indicators), further subdivided into subsections, where each indicator is assigned a specific code

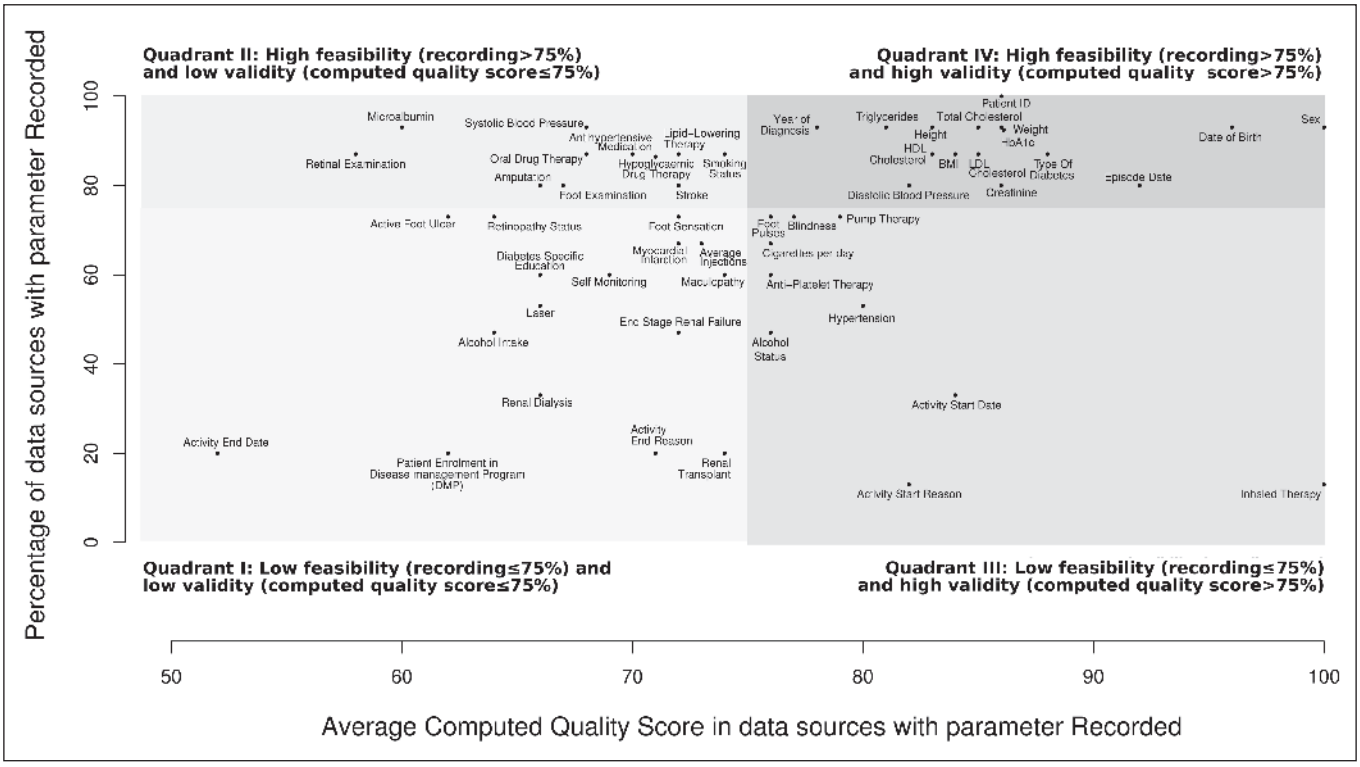

Figure 2 Bivariate distribution of BIRO data items by quality score and percentage of data sources with parameter recorded

and stratification factors (Type of Diabetes, Region, Age Class, Gender, Diabetes Duration).

A substantive extract of the final version of the EUBIROD Data Dictionary is presented in ▶Appendix 2. Several datasets were considered in the dictionary: a) the common dataset (parameters usually recorded at each encounter between a subject diagnosed with diabetes and the health care system); b) the activity dataset (indication on the status of the subject at entry/exit in the denominator within a precise time interval); c) the (general) population table (total number of residents/deaths by sex/age bands in the catchment area); d) the diabetic population table (total number of subjects with diabetes by sex/age bands in the catchment area); e) site header dataset (data custodian details); f) site profile dataset (structural characteristics of the data source and catchment area); g) the analysis dataset (derived parameters for statistical analysis e.g. age classes).

4. Discussion

The introduction of new standards can involve a number of potential consequences on existing datasets for diabetes that a recent review showed to be variously implemented within and outside Europe [34]. An expanding range of approaches is increasingly documented beyond Europe: only in the last year, major experiences were reported from the US [35], Canada [36], New Zealand [37], Malaysia [38], Singapore [39], Saudi Arabia [40]. While approaches for the development of data standards may vary, our methodology is similar to that used in the evolution of SNOMED CT [41] and HL7 [45] where experts capture all available information and evidence to harmonise and agree consistent content.

In terms of the adoption of international standards, one data binding mechanism would be certainly ideal, but not be sufficient to cover all cases of real life data collected from the many sources currently existing in Europe to monitor diabetes. We found that terminologies e.g. SNOMED CT [41] did not play a relevant role in the currently available EUBIROD databases, where heterogeneity is still a barrier to harmonization. As the adoption of electronic health record infrastructures evolves and terminologies are increasingly adopted, we may be able to extend the data dictionary by adding data item definitions using different document and data standards such as ISO 13606/openEHR [42 – 44], HL7 templates [45] or Detailed Clinical Models [46, 47]. These definitions could be used locally to set up unified data sets for data collection within national eHealth infrastructures using the provided binding mechanisms.

As far as the ISO/IEC 11179 is concerned, its constructs are "semantic" and do not describe a 'physical' or 'technical' representation of data in tables and columns. The EUBIROD data dictionary defined a technical XML representation which was directly used to implement data processing. The future adoption of ISO/IEC 11179 in the BIRO system could be beneficial to improve semantic consistency and completeness of the human readable data dictionary. The next step would be the standardisation of the technical specifications by using a standard such as SDMX [48], which could also help fostering the direct uptake of computer readable indicator definitions into existing regional or nationwide registers.

The exploitation of existing databases can provide a sustainable solution for diabetes monitoring, but requires that common rules are simultaneously applied across different data sources. This is imperative for the EUBIROD model as data are collected across several European partner sites, all using varying data standards and data collection techniques.

Building upon seminal ideas on the analysis of distributed databases in a meta-analytical fashion [49], and inspired by the practical realisation of the population-based diabetes registry of Tayside [50], the BIRO project envisaged a structured approach to ensure continuous update, consistent adaptation, and proper maintenance of standard definitions for the creation of a reliable system of health indicators. Such an approach not only improves the quality of available information, but can help clinical staff in improving daily care and enhancing the provision of complete data for secondary use, including evidence-based research and clinical audit. The EUBIROD data dictionary represents a key element of this strategy, which allowed us to respond to the main questions posed in this paper.

An essential result was that measures indicated by the scientific literature are taken into account and are routinely recorded in EUBIROD registers. The survey identified four quadrants classifying the reliability and consistency of each data item in the EUBIROD taxonomy, identifying suitable strategies for the computation of key indicators. Items falling in Quadrant IV help highlighting "gold standards" that are widely available and are consistent with the data dictionary definitions. Consequently, these items were given the highest priority for the initial development of the BIRO system.

In particular, personal data and recording dates were considered mandatory for each record entered into the system, while indicators using clinical measurements expressed as numerical data items (e.g. height, weight, cholesterol, HbA1c), which were found to be more frequent and reliable, have been extensively used to deliver statistical outputs.

Quadrants II and III included parameters at an intermediate level, identified only in a small number of datasets or with an insufficient level of detail to achieve the target definition. Smoking status is a good example of the latter, where many sources do not record a status of "ex-smoker". Blindness is defined as "Registered Blind" in several datasets, or "Receiving Money for Blindness" in DiabCare, thus reducing the consistency of data collected. In general, the two quadrants serve different purposes.

Data items at high feasibility, but low validity (Quadrant II) indicated areas where targeted strategies may improve data collection e.g. identifying a proper match between local coding and internationally agreed definitions. In the case of EUBIROD, this was particularly true for medications, examinations and risk factors/complications, as well as clinical measurements e.g. systolic blood pressure and microalbumin.

Table 3 The EUBIROD framework of diabetes indicators (ToD = type of diabetes, CAge = age classes, R = region, DS = data source, Cl. D duration = diabetes duration classes)

Section/subsection			Ref.	Indicator	Strata
1 Demographic characteristics	1.1 Basic demographics		1.1.1	Age (classes) * Gender [adult, pediatric]	DS
2 Clinical characteristics	2.1 Diabetes status		2.1.1	Type of diabetes [adult, pediatric]	CAge,DS
			2.1.2	Duration of diabetes (classes) [adult, pediatric]	ToD, gender, DS
	2.2 Risk factors for diabetes complications	2.2.1 Obesity and Growth (most recent value in the last 12 months)	2.2.1.1	Weight (classes and continuous)	ToD, gender, CAge, DS
			2.2.1.2	BMI (classes and continuous) [adult, pediatric]	ToD, gender, CAge, R
			2.2.1.3	Height (classes) [pediatric]	ToD, R
		2.2.2 Lifestyle	2.2.2.1	Smoking status	ToD, gender, CAge, R
		2.2.3 Clinical measurements (most recent value in the last 12 months)	2.2.3.1	Systolic blood pressure (classes and continuous)	ToD, gender, CAge, R
			2.2.3.2	Diastolic blood pressure (classes and continuous)	ToD, gender, CAge, R
			2.2.3.3	Total cholesterol (classes and continuous)	ToD, gender, CAge, R
			2.2.3.4	HDL-cholesterol (classes and continuous)	ToD, gender, CAge, R
			2.2.3.5	Creatinine (classes and continuous)	ToD, gender, CAge, R
			2.2.3.6	HbA1c (classes and continuous) [adult, pediatric]	ToD, gender, CAge, R
	2.3 Diabetes complications		2.3.1	Retinopathy	Cl. D duration, R
			2.3.2	End stage renal failure	Cl. D duration, R
			2.3.3	Foot ulcer	Cl. D duration, R
			2.3.4	Lower extremity amputation	Cl. D duration, R
			2.3.5	Stroke	Cl. D duration, R
			2.3.6	Myocardial infarction	Cl. D duration, R
			2.3.7	Hypertension	Cl. D duration, R
3 Health system	3.1 Structure (provider level)		3.1.1	Type of provider	
			3.1.2	Average diabetes population per region	ToD, gender, CAge
	3.2 Structural quality		3.2.1	Hospital beds per 100,000 population	R
			3.2.2	Physicians employed per 100,000 population	R
	3.3 Processes	3.3.1 Foot examination	3.3.1.1	Examination done	ToD, cage, R
		3.3.2 Eye examination	3.3.2.1	Examination done	ToD, cage, R
		3.3.3 Measurement done (in the last 12 months)	3.3.3.1	Blood pressure	ToD, cage, R
			3.3.3.2	Lipids	ToD, cage, R
			3.3.3.3	Microalbumin	ToD, cage, R
			3.3.3.4	HbA1c	ToD, Cage, R
		3.3.4 Treatment (at least one prescription in the last 12 months)	3.3.4.1	Antihypertensive medication	ToD, Cage, R
			3.3.4.2	Lipid lowering treatment	ToD, Cage, R
			3.3.4.3	Antiplatelet treatment	ToD, Cage, R
			3.3.4.4.1	Glucose lowering: diet only	ToD, Cage, R
			3.3.4.4.2	Glucose lowering: tablets only (oral drug therapy)	ToD, Cage, R
			3.3.4.4.3	Glucose lowering: insulin only	ToD, Cage, R
			3.3.4.4.4	Glucose lowering: insulin and tablets	ToD, Cage, R
			3.3.4.4.5	Glucose lowering: insulin pump	ToD, Cage, R
		3.3.5 Management	3.3.5.1	Self-monitoring	ToD, Cage, R
			3.3.5.2	Percentage with more than one visit on record	ToD, Cage, R

Table 3 Continued

Section/subsection		Ref.	Indicator	Strata
4 Population	4.1 Area level	4.1.1	Total population	Cage, R
		4.1.2	Life expectancy	Cage, R
		4.1.3	Mortality	Cage, R
5 Risk adjusted indicators*	5.1 Epidemiology	5.1.1	Prevalence of diabetes mellitus per 1,000	Region
		5.1.2	Age at diagnosis by 10-year age bands**	ToD, gender, cage, R
	5.2 Process quality (in adults with diabetes in the last 12 months)	5.2.1	Percentage with one or more HbA1c tests	ToD, R
		5.2.2	Percentage with at least one test for microalbuminuria	ToD, R
		5.2.3	Percentage with dilated eye examination or evaluation of retinal photography by trained caregiver	ToD, R
		5.2.4	Percentage receiving at least one examination of the feet	ToD, R
		5.2.5	Percentage with smoking status ascertained and documented	ToD, R
		5.2.6	Percentage with one or more serum creatinine measurements	ToD, R
		5.2.7	Percentage with one or more valid blood pressure measurements	ToD, R
		5.2.8	Percentage with hypertension who receive anti-hypertensive medication	ToD, R
		5.2.9	Percentage with oral therapy (by type)	ToD, R
		5.2.10	Percentage treated with insulin	ToD, R
		5.2.11	Percentage treated with insulin and OADs	ToD, R
		5.2.12	Percentage treated with insulin pump therapy	ToD, R
		5.2.13	Percentage with anti hypertensive treatment	ToD, R
		5.2.14	Percentage with lipid lowering treatment	ToD, R
		5.2.15	Percentage with anti platelet treatment	ToD, R
		5.2.16	Percentage performing self-monitoring of blood glucose	ToD, R
	5.3 Outcome quality: intermediate outcomes (in adults with diabetes in the last 12 months)	5.3.1	Percentage with most recent HbA1c level >9.0%	ToD, R
		5.3.2	Percentage with most recent HbA1c level >7,5%	ToD, R
		5.3.3	Percentage with most recent blood pressure less than 140/90 mmHg	ToD, R
		5.3.4	Percentage with most recent BMI > 30	ToD, R
		5.3.5	Percentage with abnormal microalbuminuria (among those tested)	ToD, R
		5.3.6	Rates of current smokers	ToD, R
		5.3.7	Rates of foot ulceration	ToD, R
	5.4 Outcome quality: terminal outcomes (in the last 12 months)	5.4.1	Annual incidence of dialysis and/or transplant in adults with diabetes	ToD, R
		5.4.2	End stage renal failure in adults with diabetes	ToD, R
		5.4.3	Annual death rate per 100,000 in the general population, for all causes and for those having diabetes as primary/secondary cause of death***	Region

· * Multivariate risk adjustment by Age, Gender, Age*Gender, Duration of Diabetes; **not adjusted; ***adjusted for standard EU population

Data items at high validity but low feasibility may imply deficiencies in the structure of local data systems, indicating areas that can only be improved long term, through targeted policies in the information infrastructure. In our study, several data items falling under this category were recognised to be inadequate for the production diabetes indicators. Notably, foot pulses, blindness and pump therapy are close borderline with Quadrant IV.

Data items included in Quadrant I were those that could not be identified in several data sources, or were formatted in ways difficult to map against EUBIROD definitions. Particular efforts are needed to resolve these issues, both in term of local policies (e.g. fostering data collection through consultation with relevant stakeholders) and in the design of more consistent standardised definitions.

Unexpectedly, key long-term complications e.g. maculopathy, myocardial infarction, retinopathy and renal complications were not well recorded by all partners. This is extremely surprising given that the original focus of the St Vincent Declaration was to "Reduce numbers of people entering end-stage diabetic renal failure by at least one third" [6], and to improve other long-term clinical outcomes.

Similarly, items related to individual practices, e.g. diabetes education, enrolment in a disease management program, self monitoring and average injections, seem to be inconsistent or inadequately integrated in diabetes registers. Hypertension as a co-morbidity diagnosis is often not recorded, despite indication of the condition from blood pressure data. Clearly there is significant room for improvement.

Targeted information policies are required to revert this situation and allow indicators to become more accessible and reliable not only for policy makers, but even for clinical teams involved in everyday practice. A system of incentives has been implemented in the UK via the Quality and Outcomes Framework [5], primarily to support GP payment, with the benefit of more complete and consistent data collection within several disease areas. Items included in Quadrant I clearly identified indicators that, albeit relevant for diabetes management, should not be considered reliable in a European report.

Overall, data items related to the activity status of the patient were either good quality but poorly recorded (start) or both low quality and poorly recorded (end). This indicates that the epidemiological concept of person at risk, which shall be applicable in population-based registers, would be very difficult to apply within the current state of the art in Europe.

The EUBIROD data dictionary and subsequent partner review provided additional insight on anomalies that can potentially affect analysis and reporting. Although EUBIROD aspires to report consistently on all of its 'core indicators', it is clear that those characterized by high consistency, quality and completeness, carry more weight and comparability than others. Such 'gold-standard' data items deserve to be emphasised when reporting on diabetes internationally.

The methodology used for the creation of the EUBIROD data dictionary can be useful in any case or context where the occurrence of the disease and the measurement of quality of care and outcomes is high priority. Compared to other approaches proposed for data dictionaries [51], ours appears to be simpler and successfully applicable to a range of settings at the international level.

The European data dictionary may represent a key resource for the EU, potentially available for further development in future initiatives. Standardised electronic data collection ensures consistency and means that the data may be converted to XML for electronic use as part of extended technologies [52].

Limitations of the approach used in our survey are that the parameter assessment criteria involved subjective judgement from the clinical reviewer and that the qualifications of those involved may not have been consistent across all sites. However, this approach was intrinsically necessary as an initial step towards the development of a common system for data analysis in order to capture local domain knowledge. The data dictionary allowed for the BIRO system to be realised and applied on real data for an objective evaluation of the quality and consistency of elements stored in diabetes registries.

It is important to clarify that the BIRO system itself does not aim to introduce a new mechanism for data collection, but is instead a way to map the existing data towards common specifications, and share through a common system only items that reach a predefined level of harmonization. The Consortia were very aware of the importance of limiting the proliferation of additional data gathering applications, which may impede data consistency and hamper clinicians through an administrative overload, even when using a user friendly tool. A separate methodological paper will explain in detail how this approach can be integrated into normal development of electronic health records.

Although clear guidelines were provided to each of the 15 registries participating to the survey, it also remains a possibility that one reviewer's assessment of quality score or consistency, for example, may not be the same as another reviewer. To minimise this effect in future, a peer review process between EUBIROD partners would allow a further level of validation. The overall rating for each parameter was consistently determined, reviewed and verified by the work package leaders at the University of Dundee.

For practical reasons, we requested only one questionnaire from each partner. However, in many countries, relevant information is stored in different silos. For instance, social insurance accounting systems may be the best source for prescribing data, while clinical datasets are preferred for all remaining aspects of diabetes care. This means that it would be more appropriate to document each individually, rather than one general questionnaire response. This revised method of data capture may be included for future revisions. Data from Slovenia and Luxembourg are sourced from paediatric data sources. This means that their clinical systems are not designed to record some of the long-term complications required by EUBIROD. In future, further general metadata must be captured on each contributing data source to capture general comments not restricted to an individual item.

A by-product of this work was the use of the standards as the basis for the first diabetes data registry for Cyprus, upon which a new model of care and uniform data collection process were defined. Hungary plans a similar approach, demonstrating wider applicability and potential for use in countries developing their information infrastructure.

5. Conclusion

The development of data dictionaries and data standards can be used to improve the quality, relevance, consistency and comparability of national information about health. In EUBIROD, a European minimum common dataset for diabetes has been created, based on an in-depth analysis of all contributing data sources. Within this dataset, European diabetes data definitions applicable with the existing clinical datasets have been documented and agreed.

To complement the dataset, a data dictionary and related technical infrastructure have been developed to capture metadata from all contributing data sources. This approach allowed embedding local knowledge from a large and expanding network into well documented, standardised XML format. The electronic storage of this data allowed adequate reference and consistent production of statistical outputs.

In EUBIROD, for the first time a diabetes data dictionary with corresponding data definitions has been created for a European population. This data dictionary forms the cornerstone for future data manipulation and presentation. When data are shared we need to ensure that all of those who need to use data can clearly understand the meaning, regardless of how the data are collected or stored at source.

In 2012, data were collected for the calendar year of 2010 from each partner site based upon the defined dataset. The results of this analysis will be published in a companion paper. Further objectives, based on suitable funding, are to expand the consortium within and beyond Europe, to analyse multiple chronic conditions using the technical infrastructure that has been already developed.

Acknowledgments

The present paper has been produced as a part of EUBIROD, a project co-funded by the European Public Health Program, DG-SANCO, European Commission, Contract no. P-2007–115.

The preparation of the manuscript has benefited from independent feedback provided by Prof. Niek Klazinga, Professor of Social Medicine, University of Amsterdam, Netherlands.

The following members of the EUBIROD consortium contributed to the development of the EUBIROD model: Gerard Boran; Adany Roza; János Sándor; Mikael Aberg; Wil Kuijpers; Andriana Evripidou; Vivie Traynor; Tadej Balletino; Iztok Stotl; Tamara Poljicanin; Sverre Sandberg; Karianne Fjeld Løvaas; Terje Tysse; Mikal Carlsen Ostensen; Philipp Perner; Stefano Gualdi; Silvia Canivell; David Whiting; Joanna Polańska; Lukasz Pracki; Carine de Beaufort.

References

1. World Health Organization. Global action plan for the prevention and control of noncommunicable diseases 2013–2020. Geneva 2013. Available at: http://apps.who.int/iris/bitstream/10665/94384/1/9789241506236_eng.pdf?ua=1 Accessed 03/06/2015.
2. Kontopantelis E, Reeves D, Valderas JM, Campbell S, Doran T. Recorded quality of primary care for patients with diabetes in England before and after the introduction of a financial incentive scheme: a longitudinal observational study. BMJ Qual Saf 2013; 22 (1): 53–64.
3. International Diabetes Federation. IDF Diabetes Atlas, 6th edn. Brussels, Belgium: International Diabetes Federation, 2013. Available at: http://www.idf.org/diabetesatlas. Accessed 03/06/2015.
4. Chen G, Khan N, Walker R, Quan H. Validating ICD coding algorithms for diabetes mellitus from administrative data. Diabetes Res Clin Pract 2010; 89 (2): 189–195.
5. Calvert M, Shankar A, McManus RJ, Lester H, Freemantle N. Effect of the quality and out-comes framework on diabetes care in the United Kingdom: retrospective cohort study. BMJ 2009. Available at: http://www.bmj.com/content/338/bmj.b1870. Accessed 03/06/2015.
6. World Health Organization (Europe) and International Diabetes Federation (Europe). Diabetes Care and Research in Europe: The Saint Vincent Declaration. Diabetic Medicine 1990; 7 (4): 360.
7. Piwernetz K. DIABCARE Quality Network in Europe – a model for quality management in chronic diseases. Int Clin Psychopharmacol 2001; 16 (Suppl 3): S5–13.
8. de Beaufort CE, Reunanen A, Raleigh V, Storms F, Kleinebreil L, Gallego R, Giorda C, Midthjell K, Jecht M, de Leeuw I, Schober E, Boran G, Tolis G. European Union diabetes indicators: fact or fiction? Eur J Public Health 2003; 13 (Suppl 3): 51–54. Available at: http://eurpub.oxfordjournals.org/content/13/suppl_1/51.full.pdf+html. Accessed 03/06/2015.
9. Final Report of the EUCID Project, 2005. Available at: http://ec.europa.eu/health/ph_projects/2005/action1/docs/action1_2005_frep_11_en.pdf. Accessed 03/06/2015.
10. DG SANCO Task Force of Major and Chronic Diseases. Major and Chronic diseases in the European Union – Report 2007. European Commission, Luxembourg, 2008. Available at: http://ec.europa.eu/health/ph_threats/non_com/docs/mcd_report_exs_en.pdf. Accessed 03/06/2015.
11. The BIRO Project official website. Available at: http://www.biro-project.eu. Accessed 03/06/2015.
12. The BIRO Consortium, Best information through regional outcomes: a shared European diabetes information system for policy and practice. Università di Perugia, Perugia, Italia, May 2009. Available at: http://www.eubirod.eu/documents/downloads/BIRO_Monograph.pdf. Accessed 03/06/2015.
13. Di Iorio CT, Carinci F, Azzopardi J, Baglioni V, Beck P, Cunningham S, Evripidou A, Leese G, Loevaas KF, Olympios G, Orsini Federici M, Pruna S, Palladino P, Skeie S, Taverner P, Traynor V, Massi Benedetti M, Privacy impact assessment in the design of transnational public health information systems: the BIRO project, Journal of Medical Ethics 2009; 35 (12): 753–761.
14. EUBIROD Consortium. EUropean Best Information through Regional Outcomes in Diabetes. Università di Perugia, Perugia, Italy. Available at: http://www.eubirod.eu/. Accessed 03/06/2015.
15. IBM Dictionary of Computing. New York, NY, USA: McGraw-Hill, Inc; 1993.
16. Australian Institute of Health and Welfare. Metadata Online Registry. Available at: http://meteor.aihw.gov.au. Accessed 03/06/2015.
17. US Agency for Healthcare Research and Quality, Health Information Knowledgebase. Available at: http://ushik.ahrq.gov/. Accessed 03/06/2015.
18. US National Cancer Institute, National Cancer Data Standards Repository. Available at: http://cbiit.nci.nih.gov/ncip/biomedical-informatics-resources/interoperability-and-semantics/metadata-and-models/. Accessed 03/06/2015.
19. The BIRO Project Consortium. WP2: Clinical Review: Indicator Development Results, January 2009. Available at: http://www.biro-project.eu/documents/downloads/D2_2_Clinical_Review_Update.pdf. Accessed 03/06/2015.
20. The BIRO Project Consortium. WP4: Data Dictionary, October 2007. Available at: http://www.biro-project.eu/documents/downloads/D4_1_Data_Dictionary_v0_2.pdf. Accessed 03/06/2015.
21. The BIRO Project Consortium. WP4: XML Metadata Dictionary, October 2007. Available at: http://www.biro-project.eu/documents/downloads/D4_2_XML_Metadata_Dictionary_v0_3.pdf. Accessed 03/06/2015.
22. The International System of Units (SI). Available at: http://www.bipm.org/en/measurement-units/. Accessed 03/06/2015.

23. Forum for Quality Systems in Diabetes Care. Available at: http://www.fqsd.at/en/. Accessed 03/06/2015.
24. Morris AD, Boyle DI, MacAlpine R, Emslie-Smith A, Jung RT, Newton RW, MacDonald TM. The diabetes audit and research in Tayside Scotland (DARTS) study: electronic record linkage to create a diabetes register. BMJ 1997; 315: 524 –528, available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2127363/pdf/9329309.pdf. Accessed 03/06/2015.
25. Massi Benedetti M, Carinci F, Orsini Federici M. The Umbria diabetes register. Diabetes Res Clin Pract 2006; 74 (Suppl 2): S200 –204.
26. Cameron AJ, Zimmet PZ, Atkins RC, Shaw JE. The Australian Diabetes, Obesity and Lifestyle Study – Profiling Diabetes and Cardiovascular Disease Risk in the Nation. Asia-Pacific Cardiology 2008; 2 (1): 25 –28. Available at: http://www.touchendocrinology.com/system/files/private/articles/8773/pdf/cameron.pdf. Accessed 03/06/2015.
27. The BIRO Project Consortium. WP4: Common Dataset, July 2009. Available at: http://www.biro-project.eu/documents/downloads/D3_1_Common_Dataset_v1_7.pdf. Accessed 03/06/2015.
28. The BIRO Project Consortium. WP4: Dictionary/XML Update, July 2009. Available at: http://www.biro-project.eu/documents/downloads/D4_3_Dictionary_XML_Update_v1_0.pdf. Accessed 03/06/2015.
29. Di Iorio CT, Carinci F, Brillante M, Azzopardi J, Beck P, Bratina N, Cunningham SG, De Beaufort C, Debacker N, Jarosz-Chobot P, Jecht M, Lindblad U, Moulton T, Metelko Ž, Nagy A, Olympios G, Pruna S, Røder M, Skeie S, Storms F, Massi Benedetti M. Cross-border flow of health information: is "privacy by design" enough? Privacy performance assessment in EUBIROD. Eur J Public Health 2012 May 4.
30. Nomenclature of territorial units for statistics (NUTS). Available at:http://ec.europa.eu/eurostat/web/nuts/overview. Accessed 03/06/2015.
31. EUBIROD Meta Registry of Diabetes Data Sources. Available at: http://www.eubirod.eu/academy/metaregistry.html. Accessed 03/06/2015.
32. Cunningham SG, Brillante M, McAlpine RR, Morris AD, Di Iorio CT, Adány R, Azzopardi J, Battelino T, Beck P, Boran G, de Beaufort C, Debacker N, Jarosz-Chobot Przemyslawa K, Jecht M, Lindblad U, Metelko Ž, Olympios G, Pruna S, Røder M, Skeie S, Storms F, Whiting D, Massi Benedetti M, Carinci F. Common Dataset. Available at: http://www.eubirod.eu/documents/downloads/D5_1_Common_Dataset.pdf. Accessed 03/06/2015.
33. R Development Core Team. R: A Language and Environment for Statistical Computing, 2010. Available at: http://cran.r-project.org/doc/manuals/refman.pdf. Accessed 03/06/2015.
34. Carinci F, Di Iorio CT, Massi Benedetti M. Standardized information exchange in diabetes: integrated registries for governance, research and clinical practice. In: Bruttomesso D, Grassi G (eds). Technological Advancements in the Treatment of Type 1 Diabetes. Front Diabetes, Basel, Karger, 2014; 24: 1–14.
35. Holt TA, Gunnarsson CL, Cload PA, Ross SD. Identification of undiagnosed diabetes and quality of diabetes care in the United States: cross-sectional study of 11.5 million primary care electronic records. CMAJ Open 2014; 2 (4): E248 –255.
36. Aronson R1, Orzech N, Ye C, Goldenberg R, Brown V. Specialist-led diabetes registries and predictors of poor glycemic control in type 2 diabetes: Insights into the functionally refractory patient from the LMC Diabetes Registry database. J Diabetes 2015.
37. Jo EC, Drury PL. Development of a Virtual Diabetes Register using Information Technology in New Zealand. Healthc Inform Res 2015; 21 (1): 49 –55.
38. Chew BH1, Lee PY, Shariff-Ghazali S, Cheong AT, Ismail M, Taher SW. Predictive Factors of follow-up Non-attendance and Mortality Among Adults with Type 2 Diabetes Mellitus – an Analysis of the Malaysian Diabetes Registry 2009. Curr Diabetes Rev 2015; 11 (2): 122–131.
39. Tan WS, Ding YY, Xia WC, Heng BH. Effects of a population-based diabetes management program in Singapore. Am J Manag Care 2014; 20 (9): e388 –398.
40. Al-Rubeaan K, Abu El-Asrar AM, Youssef AM, Subhani SN, Ahmad NA, Al-Sharqawi AH, Alguwaihes A, Alotaibi MS, Al-Ghamdi A, Ibrahim HM. Diabetic retinopathy and its risk factors in a society with a type 2 diabetes epidemic: a Saudi National Diabetes Registry-based study. Acta Ophthalmol 2015; 93 (2): e140 –147.
41. International Health Development Standards Development Organisation (IHDSDO) – SNOMED CT, 2015. Available at http://www.ihtsdo.org/snomed-ct. Accessed 03/06/2015
42. International Standardisation Organisation, ISO 13606-2: 2008, Health informatics – Electronic health record communication – Part 2: Archetype interchange specification.
43. openEHR Archetype Model, openEHR Foundation. Available at http://www.openehr.org/programs/specification/releases/currentbaseline. Accessed 03/06/2015.
44. Beale T. Archetypes constraint-based domain models for future-proof information systems, 2002. Available from: http://www.openehr.org/files/resources/publications/archetypes/archetypes_beale_oopsla_2002.pdf. Accessed 03/06/2015.
45. Health Level Seven International, HL7 Templates Standard: Specification and Use of Reusable Information, Constraint Templates, Release 1, Draft Standard for Trial Use, 2014. Available at http://www.hl7.org/implement/standards/product_brief.cfm?product_id=377. Accessed 03/06/2015.
46. Health Level Seven, Detailed Clinical Models, 2015. Availabel at http://wiki.hl7.org/index.php?title=Detailed_Clinical_Models. Accessed 03/06/2015.
47. Goossen W, Goossen-Baremans A, van der Zel M. Detailed Clinical Models: A Review. Healthc Inform Res 2010; 16 (4): 201–214. http://dx.doi.org/10.4258/hir.2010.16.4.201, PMID:23331046
48. SDMX – Statistical Data and Metadata Exchange, 2015. Available at www.sdmx.org. Accessed 03/06/2015.
49. Carinci F, Orsini Federici M, Massi Benedetti M. Diabetes registers and prevention strategies: towards an active use of health information. Diabetes Research and Clinical Practice, Volume 74, S215–S219.
50. Boyle D, Cunningham S, Sullivan F, Morris A, on behalf of the Tayside Regional Diabetes Network. Technology Integration for the provision of population-based equitable patient care: the Tayside Regional Diabetes Network: a brief description. Diabetes Nutrition and Metabolism 2001; 14 (2), 100–103.
51. de Lusignan S, Liaw ST, Michalakidis G, Jones S. Defining datasets and creating data dictionaries for quality improvement and research in chronic disease using routinely collected data: an ontology-driven approach. Inform Prim Care 2011; 19 (3): 127–134.
52. Duda SN, Cushman C, Masys DR. An XML model of an enhanced data dictionary to facilitate the exchange of pre-existing clinical research data in international studies. Stud Health Technol Inform 2007; 129 (Pt 1): 449–453.