INTERNATIONAL JOURNAL OF MEDICAL INFORMATICS 77 (2008) 58–67

journal homepage: www.intl.elsevierhealth.com/journals/ijmi

Trans-Atlantic data harmonization in the classification of medicines and dietary supplements: A challenge for epidemiologic study and clinical research

Susan Moyers, Rachel Richesson, Jeffrey Krischer*

Pediatrics Epidemiology Center, University of South Florida, College of Medicine, 3650 Spectrum Blvd., Suite 100, Tampa, FL 33612, USA

ARTICLE INFO

Article history:
Received 7 July 2006
Received in revised form
20 October 2006
Accepted 22 December 2006

Keywords:
Medications
Dietary supplements
Data standardization
Data management
Epidemiological methods
Clinical research

ABSTRACT

Objectives: As international scientific collaboration increases, there is a growing requirement for research data to be comparable among countries. Despite the importance of medication and dietary supplement data in research, there are no international standards for the collection and storage of these data. In the absence of such standards, we needed to adopt a strategy for classification and coding of medications and dietary supplements to meet demands of our multi-national study.

Methods: Given the inter-country variations in nomenclature that characterize prescription, over-the-counter (OTC) medications, traditional herbal medicines, and dietary supplements, we adopted RxNorm as a data standard for medication data, and developed an independent system that extends this standard and allows for flexible and scalable data collection for dietary supplements.

Results: RxNorm was implemented in May 2005 and as of July 2006, coverage has been 99%, at the level of active ingredients, of all the medications reported in our study. Development of a dietary supplement database began in August 2005, and has thus far coded some 1200 dietary supplements and 650 infant formula products and forms from the four countries in our study.

Conclusion: The methods we have used to collect, store, and manage medication and dietary supplement data serve as interim solutions until international standards are developed. It is hoped that such standards will ultimately emerge, and that our strategy and data model will be of value in other research environments in the immediate future.

1. Introduction

The Pediatrics Epidemiology Center (PEC), located within the University of South Florida, College of Medicine, Department of Pediatrics, acts as a Data and Technology Coordinating Center to several large multi-site and multi-national epidemiologic and clinical intervention studies, including the prospective cohort study known as TEDDY (The Environmental Determinants of Type 1 Diabetes in the Young). TEDDY [1,2] is a 15-year investigation among a population of infants and children who are genetically susceptible to Type 1 Diabetes (T1DM). The TEDDY study began enrollment in August 2005, and as of July 2006, has screened and genotyped almost 86,000 infants in the effort to find individuals with high risk

* *Corresponding author at*: Clinical Applications Unit, Mail Stop #129, 13176 North Dale Mabry Highway, Tampa, FL 33618, USA. Tel.: +1 813 948 9040; fax: +1 813 948 0010.
E-mail address: moyers@ufl.edu (S. Moyers).

1386-5056/$ – see front matter © 2007 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.ijmedinf.2006.12.003

genetic markers (HLA genotype DR 3/4, DR 4/4 and certain others). The study aim is to characterize environmental factors that help explain reasons why some genetically at-risk children develop T1DM while others do not. Investigators collect serum and plasma samples from enrolled infants beginning 90 days post-partum, then every 3 months until age 4, and every 6 months thereafter until age 15. Laboratory assays monitor for the presence of autoantibodies associated with the disease, as well as biomarkers for nutrients, cytokines, and pathogens that have been associated with either higher or lower risk of T1DM in prior research. Data collection also involves physical assessments such as growth rates, as well as numerous self-reported questionnaires about pregnancy and child diet, child and maternal medications, infections, vaccinations, and psychosocial stressors. The projected enrollment screening of over 220,000 and enrollment of over 7000 newborns involves multiple clinical centers in Europe and North America (Finland, Germany, Sweden, and the USA).

The accurate coding of medications and dietary supplements in TEDDY, as in any clinical trial, is vitally important, even though this trial, like many others, is not investigating the use of these products per se. Yet participants' use of supplements and/or medications may influence a number of important factors, such as researchers' choice of entry criteria for a clinical trial, evaluation of the effectiveness of a clinical trial, and analysis of the determinants of health and disease.

There is no single internationally recognized standard for recording and categorizing data for medications, for medical foods such as infant formulas, or for dietary supplements, and there is not even an internationally accepted definition of what constitutes a dietary supplement. Prior studies have used a variety of coding mechanisms to record data on drugs [3–6] and dietary supplements [7–9], including locally developed drug lists, as well as nationally representative commercial databases such as MediSpan, but there has not been consistency in research methodologies. Data standards at an international level are imperative not only for improved epidemiological research, but for other types of clinical research and patient care as well. This journal has served as an international forum for nations to discuss their requirements [10], challenges [11,12] and selection [10,13–17] of various healthcare data standards [13,15]. The use of standards in international settings is expensive and complicated by factors including financial, technical, cultural, legal and security and privacy issues [14]. Despite these challenges, successful multi-nation standardization endeavors have been demonstrated [18]. International data standards are the foundation for the global exchange of health information which promises to improve patient quality of life and public health [19] in coming decades, and this work moves to create discussion and consensus in our interest area of medications and dietary supplements.In an effort to produce harmonized data in our four TEDDY countries, we sought to establish both definitions and data recording standards for these constructs to enable collection of robust and comparable information on reported medications, infant formulas, and dietary supplements. We summarize our experience and highlight lessons learned and future research areas.

1.1. *Requirements for medication and dietary supplement data in TEDDY*

In the self-reported portions of data collection for the TEDDY study, subjects typically cite specific prescription and OTC medications, traditional medicines, and dietary supplements. Standardizing the collection and capture of these data is critical for meaningful analyses, but there are several "levels" of standardization to consider. A first issue is classifying the nature of the substance ingested (e.g., OTC medication, traditional medicine, dietary supplement, etc.). A second issue concerns verification of products stated in subjects' self-reports, as it is not uncommon for dietary supplements to be perceived and reported as medications, and vice versa. A third issue is determining what pieces of information should be collected for each (e.g., name, ingredients, or dosage) and whether this can be done consistently across data collection sites. The final issue is then choosing appropriate data standards for each construct and to facilitate the coding of products at the same level of detail for each of many data collection sites.

1.2. *Types of data collected and operating environment*

Data for the TEDDY study are collected by distributed research staff using scannable automated forms, and the data are transmitted electronically through scanning, or manually entered on on-line forms to the PEC. Table 1 shows items from TEDDY questionnaires that have generated subjects' responses about dietary supplements and medications.

There were three major requirements for "standard" medication and/or dietary supplement names in the TEDDY project. First, because of the use of scannable automated forms and the various languages involved, there was a need for a *code* to support each reported substance. The use of free text was not feasible because of length restrictions for the automated form fields, the high likelihood of spelling errors on forms or scanning errors due to multiple handwriting sources and multiple languages. Given the need for a alphanumeric code for each medication, the ideal standard should adhere to best practice criteria [20] regarding the management of numerical codes. The codes should be non-sensical, unique, non-ambiguous, permanent, and should never be recycled to encode other concepts. Additionally, the ideal standard for a long-term project, such as TEDDY should include a defined process to add codes as new products or gaps in the standard are discovered [21].

The second major requirement in identifying a standard for medication or dietary supplement names in the TEDDY project is that the ideal standard would be something in routine use in all of the member countries. However, as previously described, there are currently no international standard coding schemes for these products. The third requirement for a standard for is the need to store generic-named substances that are reported by study participants by brand (and generic) names. The ideal coding scheme needed to include relationships between brand names and (generic) ingredient names. To date, there are no comprehensive *electronic* sources that capture brand name—generic relationships globally, although individual countries may have reputable sources for this.

Table 1 – Study questions which elicit self-reports of medications and dietary supplements

Target subject	Question
Mother (90 days post-partum)	Did you get any vaccines during pregnancy?
	Did you take any medications during pregnancy? • Antibiotics. Please list the name of the antibiotic(s) you took • Anti-inflammatory steroid pills or injections • Medication against morning sickness • Medication for diabetes • Other?
	During your pregnancy did you take any dietary supplements such as prenatal vitamins, single vitamins, multivitamins, multiminerals, or other dietary supplements (such as fish oils, antioxidants, herbals, or others)?
Child four times yearly to age 2, and twice yearly thereafter to age 15	Has your child been given any medications—any kind of prescription medication (oral, topical, injection, etc.) and/or oral "over-the-counter" medication? • When started? (Date) • When stopped? (Date) • Reason?
	Has your child been given any dietary supplements such as single vitamins, multivitamins, multiminerals, or other dietary supplements (such as fish oils, antioxidants, herbals, or others)? • When started? (Date) • When stopped? (Date) • Type of preparation (brand name)? • How many drops, droppers, mL, or tablets did you give the child each time? • How many times per week?
	Has your child been given any vaccinations since your last TEDDY visit? • If yes, which ones?

In an effort to institute best practices for coding these products within the TEDDY study, the PEC began by identifying the three major standards requirements mentioned above. Then we evaluated against the needs of our study the status of named data standards under consideration by relevant policy and standards bodies in European Union and the United States, as well as the data collection methods of other large multi-national studies.

1.3. *Relevant standards: USA*

In the USA, the National Institutes of Health (NIH) has proposed a roadmap for research in order to establish standards for clinical trials [22]. Test projects of this roadmap include centralized data management for distributed research, the harmonization of clinical and research data, and the use of data standards throughout the research process. The Consolidated Health Informatics (CHI) initiative is a collaborative agreement between all US federal organizations that collect healthcare data. Representatives from multiple agencies have worked together to identify and recommend the use of the "best" data standards in a variety of areas (e.g., laboratory, test names, and clinical drugs) [23].

The CHI has named RxNorm as the standard for clinical drug names, principally for prescription medications. There is no named CHI standard for dietary supplements or medical foods. RxNorm is a database produced and managed by the US National Library of Medicine (NLM). RxNorm uses a controlled nomenclature of medications at varying levels of detail [24], using a simple relational model that associates multiple dosages, forms, and packaging of medications to their component ingredients (as generic or non-proprietary brand names) [25].

RxNorm does include a small number of dietary supplements; however, most of the estimated 30,000 dietary supplements sold in the USA are not part of the RxNorm inventory. A domain for dietary supplements is not even recognized by CHI, indicating that CHI will not propose standards for these products. Additionally, there is no national database for dietary supplements. A number of large research studies have developed specific supplement databases pursuant to their study needs [9,26,27]. And, a certain number of dietary supplements are included in food databases designed for nutrition research, such as the US Department of Agriculture's National Nutrient Database, and the Nutrition Data System (NDS-R) (University of Minnesota, Minneapolis, MN). NDS-R is the database used in the TEDDY study for recording food intakes. However, NDS-R does not contain information for most dietary supplements and many infant formulas reported by our subjects, and planned expansion of the dietary supplement coverage is limited.

The lack of standards and common data repository for dietary supplement information is coupled with poor visibility for researchers into a shifting and volatile product marketplace on the US side of the Atlantic. The North American dietary supplement industry experiences a lively and dynamic business environment, with continuing product reformulations, discontinuations, and new offerings in response to supply and demand. Market conditions for dietary supplements more closely align with those of the food industry, not the pharmaceutical industry. This allows for continual product change; for example, a product with the same name and manufacturer can have different ingredients and doses depending

on date of manufacture. This is a fundamental market condition that will provide enduring challenge for researchers who need to track and code products and their constituent ingredients.

The supplement markets are loosely regulated by the Dietary Supplement Health and Education Act (DSHEA), a 1994 amendment to the Federal Food, Drug and Cosmetic Act. The DSHEA does not regulate dose ranges or product constituents beyond the scope of safety and/or maximum intake limits for certain substances where adverse effects are known. Additionally, there are no regulatory measures to govern the public reporting of product changes, and, new dietary supplements require no pre-market approval unless they contain novel ingredients that were not previously sold.

At present, the NIH Office of Dietary Supplements (ODS) has made advances to create a database in conjunction with publicly downloadable SAS datasets created by the National Center for Health Statistics (NCHS). However, the NCHS data are limited to those dietary supplements actually reported by participants in the nationwide survey known as CSFII-NHANES; these datasets represent a small fraction of products being sold. Also, the NCHS datasets are compiled retrospectively, so that many entries are made obsolete either due to product discontinuance or product reformulation.

The ODS has also begun efforts to expand a proprietary industry database, so as to include label information for all dietary supplements sold in the USA. The expansion of the proprietary industry database will require the updating, coding and maintenance of an entire system of data structures, user interfaces, business logic, and other interactive components. Availability of the database is not expected until well into the next decade.

1.4. *Relevant standards: European Union*

European pharmaceutical and dietary supplement products can differ markedly from those sold in the US and may differ among the EU Member States, as do cultural practices and professional guidelines for the use of medicines and dietary supplements. And a number of factors impact product coding in the EU. Overall, the World Health Organization (WHO) has developed the International Nonproprietary Names (INN) initiative to provide a global standard naming for non-proprietary (generic) items. However, there is no accessible numerical code associated with unique drug names, or brand names, and it is the brand name that is most often reported by study participants. The INN does not maintain the many different brand names (across countries) for each generic drug. Nor is there an INN system of nomenclature for dietary supplements.

The London-based European Medicines Agency (EMEA) is the organization chartered by the EU to oversee the evaluation and supervision of medicines and to adopt a mutual recognition procedure. The current EU regulations for the naming of medications require one single name to be used for a medicinal product in all Member States. The name may be either an invented (brand) name, an INN name, or a scientific name, together with a trademark or name of the manufacturer, followed by the strength and the pharmaceutical form [28].

In a separate EU undertaking, the European eHealth Action Plan, adopted in 2004, is providing a roadmap for interoperable healthcare systems among Member States. The proposal of standards for coding medications is included in the eHealth Action Plan, and is targeting to operations that support clinicians in patient care. Among the systems under consideration are SNOMED CT [29,30] and GALEN [31–33]. SNOMED CT is a comprehensive controlled clinical vocabulary, initiated over 40 years ago under sponsorship of the College of American Pathologists. GALEN was developed as a common reference model for surgical procedures in Europe, and was later extended to include generic drugs, their composition in terms of chemicals and chemical classes, their actions, indications, and interactions, based on a language independent model.

Product standards for dietary supplements sold in the EU are just evolving under the auspices of two recent regulatory documents, the Food Supplements Directive (FSD) [34] and the Traditional Herbal Medicinal Products Directive (THMPD) [35]. The THMPD, adopted in 2004, created a group within the EMEA, the Committee on Herbal Medicinal Products, charged with establishing product registrations, monographs, and indications for substances such as herbs, roots, algae and other natural products which are used as folk or traditional medicines. At present, the Committee has scheduled regular and frequent meetings in an effort to create a list of authorized products and doses. But the creation of the approved substances/doses list is a major undertaking, and it is unclear when the list will be made available.

The FSD, initially adopted in 2002, creates a single market in non-medicinal food supplements throughout the EU by integrating regulation for ingredients, forms of ingredients, and doses. The FSD directed an extensive safety evaluation, followed by regulations for marketing and labeling food supplements. Currently, the FSD deals with only vitamins and minerals and will not go into full effect until December 31, 2009. The FSD is likely to be extended to cover additional products, possibly including some herb and plant extracts that are not listed under THMPD. Additionally, the FSD allows for local country regulations to prevail in cases where country-specific regulations differ from those specified under the FSD.

2. Approach used in our study

An early challenge for TEDDY was the ascertainment of specific definitions to use for the various products, as the distinction between medication and dietary supplement is often blurred. We considered the following classifications:

(1) *Dietary supplements*: There are no internationally accepted definitions. In the USA, DSHEA provides that dietary supplements can be vitamins, minerals, herbals and botanicals, animal extracts, amino acids, proteins, concentrates, metabolites and constituents, teas, or other miscellaneous products. Indeed, all that is needed for a product to qualify as a dietary supplement in the USA are the words “dietary supplement” on the product label and compliance with other product labeling regulations. In the EU, the FSD (2002/46/EC) defines food supplements

as "foodstuffs the purpose of which is to supplement the normal diet and which are concentrated sources of nutrients or other substances with a nutritional or physiological function, marketed in dose form." At present, FSD applies to a restricted list of vitamins and minerals, but in time could apply to all products that contain established nutrients. The THMPD defines traditional medicines as products sold over-the-counter that have a history of use anywhere in the world for at least 30 years, and in the EU for at least 15 years. They may be herbs, extracts, or other types of products.

(2) *Medications*: Medications in the USA include those products that meet the definition of drugs in Section 201 of the Federal Food, Drug and Cosmetic Act (21 U.S.C. 321) [36], intended for use in the diagnosis, cure, mitigation, treatment, or prevention of disease, including those that are chemically synthesized and those derived from living sources (biologic products). The definition encompasses any product recognized in the official United States Pharmacopeia, official Homeopathic Pharmacopeia of the United States, official National Formulary, or any supplement to any of them. In the EU, Directive 2004/27/EC of 31 March 2004 [37] defines as medicinal any substance or combination of substances presented as having properties for treating or preventing disease in human beings; or any substance or combination of substances which may be used in or administered to human beings either with a view to restoring, correcting or modifying physiological functions by exerting a pharmacological, immunological or metabolic action, or to making a medical diagnosis.

2.1. *Operational considerations*

The blurred boundaries among prescription and other medicines, and dietary supplements are often exacerbated by difficulties in obtaining complete and accurate data from subject self-report. To collect the most accurate and detailed information, staff members in the TEDDY study ask participants to bring to the clinic the product labels and bottles, but typically the subjects do not. And, a large percentage of study participants are unable to identify their medications and dietary supplements beyond the brand name, or general product description. Research staff thus must engage in the often tedious and time-intensive process of determining product ingredients and dosage, then making a determination about how to define and code them within the study's data structures.

2.2. *Standard rationale and implementation—RxNorm for medications*

RxNorm was selected as the ideal standard for coding medications in the TEDDY study because it meets all of the requirements outlined above. While it is not international in mission, the NLM does strive to address international needs in all of its activities, and as a rule ensures free and global access to its products. The library guarantees "best" terminology practices [20], and the numerical codes in RxNorm will never be changed or reassigned. These code numbers are easily accessible through the NLM-developed UMLS and the RxNav [38] interfaces.

RxNorm uses a controlled nomenclature of medications at varying levels of detail [24], using a simple relational model that associates multiple dosages, forms, and packaging of medications to their component ingredients (as generic or non-proprietary brand names) [25] (Fig. 1.) RxNorm provides standard names for clinical drugs (active ingredient + strength + dose form) as administered to patients. It provides links from clinical drugs, both branded and generic, to their active ingredients, drug components (active ingredient + strength), and related brand names.

In the USA, National Drug Codes (NDCs) adopted by the FDA for specific products (where there are often many NDC codes for a single product) are linked to a specific item in RxNorm. RxNorm also links its names to medication lists commonly used in pharmacy management and drug interaction software. The current focus of RxNorm is on US drugs to support the representation of administered drug products for hospital information systems [24] RxNorm contains prescription medications and some OTC products (e.g., single vitamin preparations, pain relievers, antacids).

2.3. *Standard rationale and implementation—dietary supplements*

Given the lack of standard definitions for dietary supplements and traditional medicines, the TEDDY study decided to generally implement the DSHEA definition, because it is wider in scope and includes products covered by both the FSD and THMPD in the EU. Study personnel follow procedures to record any pill, capsule, extract, or powder as a dietary supplement unless it is identifiable through RxNorm as a medication. A vitamin or other nutrient that is not administered orally (such as intramuscular or subcutaneous injection) is recorded as a medication. Also, if an herbal product or other DHSEA-defined dietary supplement is taken on the advice of a health care practitioner in connection with treating a health condition, it is recorded as a medication, in keeping with the language of the THMPD. Additionally, homeopathic products are recorded as medications because, while often confused with dietary supplements, they are licensed as medicines in the TEDDY countries where they are sold, and are included in the definition of medicines in the US Food, Drug and Cosmetic Act.

In our efforts to identify a suitable data repository for dietary supplements and traditional medicines taken in the absence of recommendation by a health professional, we concluded that no existing coding or data collection system meets the need of a large-scale multi-national trial such as TEDDY. The decision was made to create a relational database to catalog products reported by our subjects. The structure of the dietary supplement database was inspired by the RxNorm data model and is scheduled to be under continuous update as new products are reported throughout the study. The model captures relationships between brand name, ingredient, ingredient strength, and product form, and enable a user to query from the specific to the generic and vice versa, to allow data entry and look up by varying levels of detail (e.g., brand, ingredient, dosage, form, packaging,) with defined relationships between these constructs, as shown in Fig. 2. This

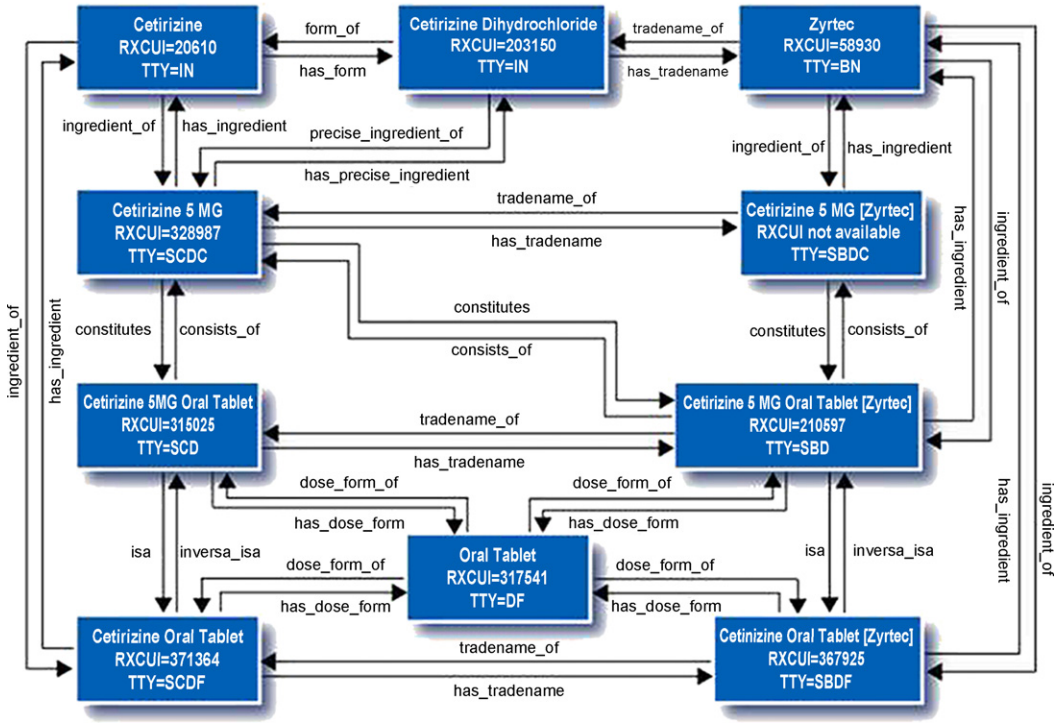

Fig. 1 – RxNorm data structures. Source: http://www.nlm.nih.gov/research/umls/rxnorm/overview.html

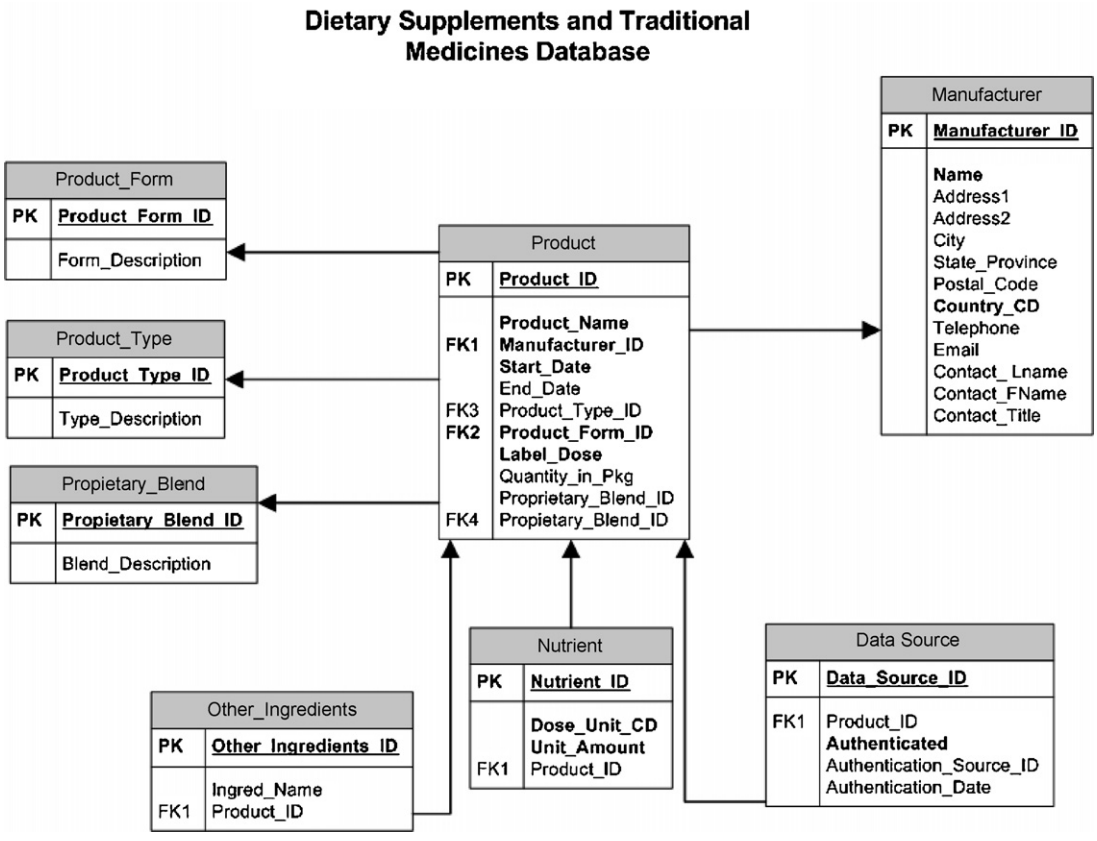

Fig. 2 – Data model.

structure therefore allows data to be recorded at varying levels of detail, but the relationships between the levels of detail are explicit.

Products are keyed by brand name, as this is the way study participants typically report them. In addition to brand name, the database includes variables to capture manufacturer, dosage, dosage form, and the ingredient content of 40 constituents and their chemical form. For products that are discontinued or reformulated, PEC staff obtain the entry and exit date of each formulation and record that data into separate variables. There is a variable to capture the source of our information (product label, supplier, ODS, etc.). Another variable captures miscellaneous product attributes, such as filler and binder constituents, and substances not recognized as containing nutrients, such as many herbs. We have also included variables to record whether the product dose and ingredients have been authenticated in laboratory assays, and the source of that authentication. This information is needed due to lingering issues regarding manufacturing quality control; some dietary supplements have been found to contain ingredients and/or doses substantially different than those listed on the label.

In addition to DSHEA-defined dietary supplements, our database includes other products, such as infant feeding formulas. While we found that the ingredients of many infant formulas are available in our four country-specific food databases, there is no data repository national or otherwise, to capture data at the level of detail desired for the TEDDY study. For example, many infant formulas now include functional components such as prebiotics/probiotics, or include specialized or hydrolyzed proteins—constituents which are not readily identifiable in traditional food databases, but may be valuable in the analyses of outcomes. The dietary supplements database also records certain other traditional medicines and functional foods reported by our population (such as vitamin waters, and highly fortified nutritional drinks).

2.4. Process for collecting medication data in TEDDY project

The process identified for the coding of medications in TEDDY includes reporting medications by generic-named ingredients to the PEC, where the RxNorm codes are centrally assigned and managed for the project. If the reported medication has been collected previously in the TEDDY study, it is included in an on-line codebook, where research staff can access the correct code. If the medication name has not been previously collected, field researchers contact the PEC, whose staff determines the correct RxNorm code using the NLM's RxNav tool [38]. Because there is no single source for international brand name—generic name, each TEDDY site is responsible to determine generic (active) ingredients from any brand name reported medications from best national sources (when the brand name cannot be found in RxNorm). [While the World Health Organization's International Nonproprietary Names (INN) provide a well-adopted global standard naming for nonproprietary (generic) names, there is no accessible numerical code associated with unique drug names. Additionally, the INN does not maintain data on the many different brand names (across countries) for each generic drug.] The PEC developed an on-line codebook (linking multiple proprietary names to unique generic ingredients or unique combinations of generic ingredients) to facilitate users to access a subset of (common) RxNorm codes remotely and on demand. Further, the PEC has an internal system in place to deal with generic drug names that might not be found in RxNorm. This process includes the assignment of a temporary ID number while the PEC contacts NLM about including the medication in RxNorm.

2.5. Process for collecting dietary supplement data in TEDDY project

The process identified for the coding of dietary supplements is similar to that for medications. All information is collected in a standardized way by centrally trained and certified local interviewers, supervised by centrally trained and certified site nutritionists, the country lead nutritionist, and the international nutrition manager. During the TEDDY clinic interviews, if the product and/or label are not provided by the subject, clinical staff make extensive probes to ascertain the product, product form, dose taken, and related information. Following the clinic interview, clinical staff review existing codes of products that have already been reported and associated with the scannable forms. For any product not already coded, the site staff, together with the PEC, assign a new code, followed by identification of label doses, dose forms, and ingredients. This typically involves contacting the manufacturer or distributor directly by telephone or electronic mail.

2.6. Status

RxNorm was implemented in the TEDDY project in May 2005 and as of July 2006, coverage has been 99% (282/284), at the level of active ingredients, of all the medications reported in the TEDDY study. Development of the dietary supplement database began in August 2005, and has thus far coded some 1200 dietary supplements and 650 infant formula products and forms from the four countries in our study. We have learned that pediatric supplements in particular include nontraditional and creative dose forms—fizzies, gummies, pops, and droppers, among others, in addition to the customary tablets, capsules and softgels, and thus we have expanded the entries in dose forms variable. Additionally, we found it helpful to distinguish products by country because two or more countries can market products with same or similar names, but different ingredients or doses.

The definitions and operational procedures we have adopted for medications and dietary supplements provide guidance for researchers to collect data systematically and to store it in specific places in the TEDDY data repository. The characteristics of medications and dietary supplements imply different data standards. These data are collected in different parts of the data collection forms, so that single data fields are encoded in a single data standard.

The models for both medications and dietary supplement data are similar, and both allow for the storage, retrieval and analysis of data at varying levels of specificity. Until there are named international or trans-Atlantic data standards for the collection and storage of medication and dietary supple-

ment data, this solution allows TEDDY researchers to collect, manage, and analyze data to meet study objectives. Our implementation of a data standard and standardized data collection practices has implications to benefit not only epidemiological research but also applications in patient safety [6] that require standard coded medication information.

Summary points

What was known before the study?

- Epidemiologic and other human research studies collect data on the use of pharmaceutical products and dietary supplements.
- There is no standard international or global coding or classification scheme for these data.
- There is no standard practice for collecting and storing these data, nor is there a standard data model to guide local efforts.

What this study has added to the body of knowledge?

- This work reviews current data standards and "best" practices (both in the USA and the EU) for the collection and storage of data related to the use of pharmaceutical products and dietary supplements.
- This work highlights the challenges in defining international data standards and "best" practices for the collection and storage of these data, and highlights differences in data standards policy in the USA and EU.
- In lieu of commonly accepted data standards, a model for storing reported pharmaceutical and dietary supplement data is presented. This data model and the described data collection practices (used in a large multi-national epidemiologic study) fulfill epidemiologic research data requirements and have the potential to convert to future standards when they are established.

3. Discussion

We recognize the immediate need for international data standards in the area of medication and dietary supplement data. In lieu of accepted standards, we have developed an option that captures standards requirements for epidemiologic research, and has the potential to convert to future standards when they are named. Although RxNorm is solely a US standard, at the ingredient level, most medications reported on the TEDDY study to date have been described in RxNorm. The use of the standards included in RxNorm can facilitate the sharing and exchange of research data across studies as has been called for by the NIH [39]. A standard that represents the relationship between the different granularities of data collection across countries was required. For later analysis and data mining activities, a desirable feature of a data standard would be one with explicit relationships between the different granularities of medication data (e.g., "Acetaminophen" is related to "Acetaminophen 100 mg" is related to "Acetaminophen 100 mg Extended Release Tablet"). The ideal representation for medication data needs a structure that will allow unique codes for these varying levels of detail (since information is often reported to research staff at different levels of specificity), yet capture the relationships between them. Therefore, researchers can use different tables from the relational RxNorm model to capture information across countries, but relationships between the data of differing levels of specificity are explicit and can later be exploited. This same rationale supported the development of the model for storing dietary supplement data. It is critical in outcomes analysis to aim for completeness and accuracy in the collection of data regarding traditional medicines, infant formulas, and dietary supplements. For example, most dietary supplements contain substances known as "micronutrients" in the nutrition vocabulary; that is, they are nutrients needed only in small quantities to support life and growth. But micronutrients are often found to be highly levered—small changes in intakes have the potential to result in large differences in the determination of health and disease. Therefore the accurate and detailed representation of these substances in our datasets, alone, in combination with one another, and in the context of total diet including food, is necessary. We hope that the efforts of ODS to create and support USA data standards for dietary supplements continue, and we hope that emerging standards across the Atlantic address the same requirements that motivated our solution. The National Library of Medicine's UMLS System [40–42] is a source of linking multiple (over 100) standard healthcare standard vocabularies and could be tool to link RxNorm to SNOMED CT or other classifications that emerge as standards in this area.

Because of the planned longevity of the TEDDY study, the space of medications and dietary supplements that could conceivably be reported over the study life is impossible to define, as TEDDY is targeted to collect data until at least the year 2020 and beyond. The tools we have adopted are flexible and allow distributed research staff to search the entire data space, including generic and proprietary names. We recognize that standards for nomenclature, coding, and data harmonization are evolving, and we have structured our designs to have flexibility to accommodate emerging standards. Our prime objective was to create data structures and coding methodologies that are robust, detailed and operate in a platform-independent software environment that can serve as efficient interim repositories until global, or at least trans-Atlantic standards are adopted. The data structures and data collection methodologies set in place for the TEDDY study will hopefully facilitate migration once international standards for medications and dietary supplements emerge, and common, publicly accessible databases for coding these products are realized.

4. Conclusion

Clinical researchers require data collection and coding systems that can facilitate the rapid capture, coding, and catego-

rization of products, according to brand name and ingredients, adverse effects, with perhaps the added capability in the case of dietary supplements and traditional medicines, to include a grading system for authentication of product constituents, doses, and drug-supplement interactions. Data standards requirements for a large epidemiological study led to the selection of RxNorm as an ideal standard for medication data, and promoted the design of new data structures to serve as interim repository for dietary supplement data, pending evolution of trans-Atlantic or global product and coding standards.

Acknowledgements

This research was funded by grant number DK63970 from the US National Institute of Diabetes, Digestive and Kidney Diseases (NIDDK). We are grateful for the assistance of Stuart J. Nelson, MD, head of the Medical Subject Headings Section at the US National Library of Medicine for inputs regarding RxNorm; to Jennifer Hudson for data entry and coding to Jamie Malloy for efforts toward building a system to support data collection; to Wendy McLeod and Cristina McCarthy for assigning and verifying codes; and to Dulce Garcia for her assistance developing the data model.

REFERENCES

[1] PEC, TEDDY Project Website, Pediatrics Epidemiology Center, 2006.
[2] W.A. Hagopian, A. Lernmark, et al., TEDDY – the environmental determinants of diabetes in the young: an observational clinical trial, Ann. N. Y. Acad. Sci. 1079 (2006) 320–326.
[3] A.R. Keefe, et al., A hierarchical approach to coding chemical, biological and pharmaceutical substances, J. Toxicol. Environ. Health Part B: Crit. Rev. 8 (3–5) (2005) 145–452.
[4] F.B.P.P. Investigators, The Family Blood Pressure Program (FBPP): multi-center genetic study of hypertension, Hypertension 39 (2002) 3–9.
[5] W.M. Thomson, A medication capture and analysis system for use in epidemiology, Drugs Aging 10 (4) (1997) 290–298.
[6] H. Miillerovh, J. VlCek, Drug information centre-analysis of activities of a regional centre, Int. J. Med. lnform. 45 (1997) 53–58.
[7] K. Radimer, et al., Dietary supplement use by US adults: data from the National Health and Nutrition Examination Survey, 1999–2000, Am. J. Epidemiol. 160 (4) (2004) 339–349.
[8] J.A. Foote, et al., Factors associated with dietary supplement use among health adults of five ethnicities, Am. J. Epidemiol. 157 (10) (2003) 888–896.
[9] V. Newman, et al., Dietary supplement use by women at risk for breast cancer recurrence, J. Am. Diet. Assoc. 98 (3) (1998) 285–292.
[10] P.L. Elkin, et al., Guideline and quality indicators for development, purchase and use of controlled health vocabularies, Int. J. Med. Inform. 68 (1–3) (2002) 175–186.
[11] P. Williams, How to create awareness and ensure broad dissemination of health informatics standards, Int. J. Med. lnform. 48 (1998) 49–56.
[12] J. Dudeck, Aspects of implementing and harmonizing healthcare communication standards, Int. J. Med. lnform. 48 (1998) 163–171.
[13] D.E. Bourne, L.A. Hanmer, K.J. Heavens, Towards developing a national health information standards framework for South Africa, Int. J. Med. lnform. 48 (1998) 29–31.
[14] Y. Zhang, Y. Xu, L. Shang, K. Rao. An investigation into health informatics and related standards in China, Int. J. Med. Inform. 2006 June 20, in press.
[15] R. Lenz, M. Beyer, K.A. Kuhn. Semantic integration in healthcare networks, Int. J. Med. Inform. 2006 June 10, in press.
[16] G.O. Klein, Standardization strategy from a European perspective, Int. J. Med. lnform. 48 (1998) 67–70.
[17] R.C. Alvarez, J. Zelmer, Standardization in health informatics in Canada, Int. J. Med. lnform. 48 (1–3) (1998) 13–15.
[18] C. Zywietz, SCP-ECG and vital signs information representation–two examples of successful transcontinental cooperation in medical informatics standardization, Int. J. Med. lnform. 48 (1998) 195–199.
[19] P. Treseder, P. Williams, The common principles of health informatics standardisation that require exchange of information between the standardisation bodies of different countries, Int. J. Med. lnform. 48 (1998) 39–42.
[20] J.J. Cimino, Desiderata for controlled medical vocabularies in the twenty-first century, Methods Inform. Med. 37 (4–5) (1998) 394–403.
[21] M.S. Tuttle, S.H. Brown, K.E. Campbell, J.S. Carter, The semantic web as "Perfection Seeking": a view from drug terminology, in: I. Cruz, et al. (Eds.), The Emerging Semantic Web: Selected Papers from the First Semantic Web Working Symposium, IOS press, Amsterdam, 2002.
[22] NIH, NIH Roadmap, Accelerating Medical Discovery to Improve Health, National Institutes of Health, 2005.
[23] CHI, CHI Executive Summaries, Consolidated Health Informatics, 2004.
[24] NLM, RxNorm, UMLS RxNorm description and links, 2005.
[25] S. Liu, W. Ma, R. Moore, V. Ganesan, S. Nelson, RxNorm: prescription for electronic drug information exchange, IT professional 7 (5) (2005) 17–23.
[26] S.L. Archer, et al., Association of dietary supplement use with specific micronutrient intakes among middle-aged American men and women: the INTERMAP Study, J. Am. Diet. Assoc. 105 (7) (2005) 1106–1114.
[27] E.A. Yetley, Y.K. Park, Obtaining data on intake of supplements, in: L. GE (Ed.), Nutritional Status Assessment of the Individual, Food & Nutrition Press, Trumbull, CT, 1989, pp. 113–123.
[28] EMEA Committee for Human Products, Guideline on the Acceptability of Invented Names for Human Medicinal Products Processed through the Centralized Procedure, CPMP/328/98, European Medicines Agency, Human Medicines Evaluation Unit, 2005.
[29] K.A. Spackman, SNOMED CT milestones: endorsements are added to already-impressive standards credentials, Health Inform. 21 (9) (2004), p. 54, 56.
[30] J.M. Kim, P. Frosdick, Description of a drug hierarchy in a concept-based reference terminology, Proc. AMIA Symp. (2001) 314–318.
[31] A. Rossi Mori, F. Consorti, Exploiting the terminological approach from CEN/TC251 and GALEN to support semantic interoperability of healthcare record systems, Int. J. Med. Inform. 48 (1–3) (1998) 111–124.
[32] W.D. Solomon, et al., A reference terminology for drugs, in: Annual Fall Symposium of American Medical Informatics Association, Washington, DC, Hanley & Belfus Inc., Philadelphia, 1999.
[33] J. Rogers, A. Roberts, D. Solomon, et al., GALEN ten years on: tasks and supporting tools, in: R. Haux, et al. (Eds.), MedInfo 2001 Studies in Health Technology and Informatics, IOS Press, Amsterdam, 2001, pp. 256–260.

[34] EU, Directive 2002/46/EC of the European Parliament and of the Council of 10 June 2002 on the Approximation of the Laws of the Member States Relating to Food Supplements (Text with EEA relevance), Official Journal L 183, 12/07/2002 P., E.P.a.o.t. Council, 2002, pp. 0051–0057.
[35] EU, Directive 2004/24/EC of the European Parliament and of the Council of 31 March 2004 amending, as regards traditional herbal medicinal products, Directive 2001/83/EC on the Community Code Relating to Medicinal Products for Human Use, Official Journal L 136, 30/04/2004, European Parliament, 2004, pp. 0085–0090.
[36] FDA, [21 U.S.C. 301], The Federal Food, Drug, and Cosmetic Act, U.S.D.o.H.a.H. Services, U.S. Food and Drug Administration.
[37] EU, Directive of 2004/27/EC of the European Parliament and of the Council of 31 March 2004 amending directive 2001/83/EC, Official Journal of the European Union, 2004.
[38] NLM, RxNav, A Semantic Navigation Tool for Clinical Drugs, National Library of Medicine, 2005.
[39] NIH, Final NIH Statement on Sharing Research Data, February 26, 2003, N.I.o. Health, 2003.
[40] O. Bodenreider, et al., Evaluation of the unified medical language system as a medical knowledge source, J. Am. Med. Inform. Assoc. 5 (1) (1998) 76–87.
[41] NLM, Fact Sheet, UMLS® Metathesaurus, National Library of Medicine, 2003.
[42] S.J. Nelson, T. Powell, B.L. Humphreys, The unified medical language system project, Int. J. Med. lnform. (2001) 91–105.