

REVIEW

Will Systematized Nomenclature of Medicine–Clinical Terms improve our understanding of the disease burden posed by allergic disorders?

C. R. Simpson^{*}, C. Anandan[†], C. Fischbacher[‡], K. Lefevre[§] and A. Sheikh[†]

^{*}Department of General Practice and Primary Care, Primary Care Clinical Informatics Unit, The University of Aberdeen, Scotland, UK, [†]Allergy & Respiratory Research Group, Division of Community Health Sciences, General Practice Section, University of Edinburgh, Scotland, UK, [‡]Information Services Division, NHS National Services Scotland, Karen Lefevre, Scotland, UK, and [§]Department of General Practice and Primary Care, The University of Aberdeen, Scotland, UK

Clinical and Experimental Allergy

Summary

Analysis of data collected through the use of high-quality computerized systems is vital if we are to understand the health burden from allergic disease. Coding systems currently used, such as the World Health Organization's International Classification of Diseases and the Read system, have however been criticized as being unduly restrictive and hence inadequate for the detailed coding of allergic problems. Greater granularity of coding can be achieved by using the Systematized Nomenclature of Medicine–Clinical Terms (SNOMED-CT) system, which will be adopted by several countries including the United States and United Kingdom. Before the introduction of SNOMED-CT, it is important that several issues are resolved, including ensuring that adequate mapping occurs from existing systems, that the SNOMED-CT is trialled before general implementation, and that training is provided for users new to coding as part of their clinical practice. Of particular importance is that the allergy fraternity bring to light any gaps in allergy coding through the creation of a working group to advise the newly formed International Healthcare Terminology Standards Development Organisation. There is also a role for allergy experts, working in conjunction with government agencies and professional bodies, to determine a recommended set of codes, which will obviate some of the inevitable challenges raised by a very fluid coding structure for those wishing to undertake secondary analysis of health care datasets.

Correspondence:

Aziz Sheikh, Allergy & Respiratory Research Group, Division of Community Health Sciences, General Practice Section, University of Edinburgh, 20 West Richmond St, Edinburgh EH8 9DX, UK
E-mail: aziz.sheikh@ed.ac.uk

Introduction

Important gaps exist in the information available to policy makers when attempting to determine the disease burden from allergic disorders [1]. Such gaps can increasingly be filled by greater use of data that are routinely collected through the use of high-quality computerized systems in primary and secondary care, where a great wealth of patient-specific medical data are now stored. These computerized systems utilize electronic patient records, which allow users to be supported through the provision of complete and accurate data, alerts, reminders and clinical decision support systems. Data linkage techniques are increasingly being used to combine data from a variety of medical sources [2]. However, one important limitation of using these data sources is a lack of a standardized coding system, thus rendering the analyses and interpretation of these data complex.

Models of health and disease are constructed around medical coding 'languages' and this has created a need for mechanisms to collect, analyse and interpret clinical data. This need has given rise to a process of coding, whereby a set of words describing a medical concept are translated into codes for later analysis [3]. Over time, several medical coding systems have emerged (see Table 1). The use of these coding systems has varied between user populations (discussed below), with terms used in some coding systems being useful for generalists, but not necessarily as useful for specialists who may interpret the same terms somewhat differently [4].

Our aim is to describe the coding nomenclatures available within the Scottish health care system for gauging the health burden of allergic disease, and set out the advantages and disadvantages of these and other international coding systems. We also aim to set out recommendations for the international research and clinical community before the introduction of the Systematized

Table 1. Examples of coding systems

Clinical term	Description	Advantages	Disadvantages
Clinical Terms Version 3	Composed of several linked hierarchical structures with unlimited depth	Supports the construction of more detailed concepts using a mechanism of qualifiers. Has a defined syntax for permissible combinations	Only employs a limited compositional scheme
International Classification of Primary Care-2	Uses alphanumeric codes based on body systems	Coding chapters based on body systems and tailored to provide family practitioners a clinically useful standard system of coding	Simple classification system lacking in granularity with only 1371 terms
International Classification of Disease-10	Captures morbidity and mortality figures in a standard alphanumeric format. The number of classification categories doubled to 8000 with the introduction of version 10 which has 21 chapters	Allows comparisons over time and between countries	Simple classification scheme – not suitable for use in complex clinical records
Office of Population and Census Surveys Classification of Surgical Operations and Procedures (OPCS4)	Used as the primary classification for surgical operations and procedures	Contains 6000 codes with 23 separate chapters which deal with a different system of the body. Where more precision is required, a subsidiary code exists to indicate surgical methods used and specific sites of operation	Simple classification scheme – not suitable for use in complex clinical records
Read (5-byte version 2)	Hierarchical structure with unambiguous five character alpha-numeric codes that have preferential and synonymous clinical terms	Hierarchical design allows coding to the highest level of certainty	Limited number of concepts available therefore great wealth of information entered as free text
Systematized Nomenclature of Medicine-Clinical Terms	A compositional terminology which is hoped will enable healthcare professionals to communicate in clear and unambiguous terms	User can create own sets of codes useful to clinical practice (post-coordinated concepts) whereby concepts are built up from other concepts and their codes. Allows data transfer between countries and systems utilising this coding system. Allows detailed coding without compromising fluidity	Could complicate the ability for secondary analysis. Users will need in depth knowledge. The system has been hitherto untried

Nomenclature of Medicine-Clinical Terms (SNOMED-CT) system.

International Classification of Diseases

Maintained by the World Health Organization, the International Statistical Classification of Diseases and Related Health Problems (ICD) was by 2001 in its 10th revision (ICD-10). Its history can be traced back to the beginnings of medical terminology when, in 1895, William Farr and colleagues developed an International List of Causes of Death [4]. The goal of the ICD-10 is to allow for morbidity and mortality data to be systematically and consistently recorded allowing comparisons over time and between countries. ICD-10 is a single list of three alphanumeric character codes ranging

from A00 to Z99. For more detail, a fourth numeric character can be added after a decimal point [3]. The 10th revision classifies diagnoses, signs and symptoms, causes and other factors into 21 alphanumeric chapters [4].

Office of Population and Census Surveys Classification of Surgical Operations and Procedures

The Office of Population and Census Surveys Classification of Surgical Operations and Procedures (OPCS) version 4 coding system is used as the primary classification for surgical operations and procedures and contains about 6000 codes. These codes consist of 24 separate chapters, which deal with different systems of the body, diagnostic procedures and miscellaneous operations. Where more

precision is required, a subsidiary code exists to indicate surgical methods used and specific sites of operation [5].

Read Clinical Classification or 'Read codes'

Read was produced for clinicians in primary care for the coding of events stored in electronic patient records. Version 1 was compact in nature, as early computer systems had little memory. Four-byte Read (version 2) was completed in the mid-1980s and aimed to expand the codes required for clinical practice [6]. This coding system was expanded once more to cover secondary care with the introduction of 5-byte Read codes (version 2) in 1990 [6]. The Read nomenclature is now used by the majority of primary care electronic patient record systems in the United Kingdom including GPASS, EMIS and In Practice Systems (InPS) Ltd [4, 6]. Each preferred term in 5-byte Read (version 2) has an unambiguous five character alphanumeric code (of which 125 000 exist) that gives a unique identifier to preferential (a preferred way of describing a specific concept) and synonymous (a different way of describing the preferred term) clinical terms. Concepts are arranged hierarchically, with the level of detail increasing down the hierarchy. This hierarchical design allows the coding of a disease, symptom, procedure or occupation to a high level of certainty or detail and is intended to help users to find related terms [3].

International Classification for Primary Care

The International Classification for Primary Care (ICPC), currently in its second version, was developed by the World Organization of Family Doctors (WONCA) and has been translated into 20 languages and operates in a large number of countries [6]. This alphanumeric coding system is based on body systems and is tailored to provide family practitioners with a standard system of coding, which can be integrated into the electronic medical record. This system is comprised of 1371 terms and brings together symptoms and reasons for contact, diagnoses, interventions and operates a logical structure of data in the frame of stand alone or sequential consultations.

Limitations of existing coding systems in relation to allergy coding

A key disadvantage of the ICD-10 system is that this classification system was designed for tabulating and reporting disease and therefore has some inherent limitations of forcing artificial categorizations that are not clinically meaningful [7].

The existence of abbreviations in Read codes, which evolved from standard classifications such as ICD-10 and the OPCS classification of surgical operations, is also problematic. For example, the 'not otherwise specified'

(NOS) abbreviation is the most commonly used of these abbreviations. Thus, if a clinical presentation has been recorded in the paper clinical record as simply 'eczema', the eczema NOS (M12z1) code will sometimes be used rather than a more descriptive code such as erythrodermic eczema (M12z4). Clinical staff also have a tendency to use a more generic parent code further up the hierarchy such as (M12..) eczema [8].

It is a fundamental requirement of most clinical systems to map clinical information such as diagnoses into codes, therefore a standard set of these codes is useful for interrogating clinical databases [9]. Exhaustive sets of ICD-10 and 5-byte Read (version 2) codes have been produced to investigate the health burden of allergic disease in primary care (see Tables 2 and 3). Other sets of codes have been created for feedback reports (workload, patient management, etc.) in primary care practices and to provide information on allergy to drugs for an emergency care summary (ECS) (a Scottish wide initiative for the transfer of information from primary care records to out-of-hours services) [10].

We have found several limitations in the creation of an exhaustive list of allergy codes for use in interrogating a Scottish secondary care and general practice databases (held by the Information Services Division of the National Services Scotland and the University of Aberdeen's Primary Care Clinical Informatics Unit) (see Tables 2 and 3). These problems have largely been caused by the limited opportunity for detail offered by ICD-10 and 5-byte Read (version 2). For ICD-10, very few concepts were available (48 concepts). Although comparatively more allergy concepts were found for Read codes (223 concepts), there was a paucity of food allergy concepts, with only four specific concepts related to egg (SN580 and SN581) and nut (SN582 and SN583) allergy and three more general concepts related to food (14M1.), plant (14M2.) and multiple (14M3) allergies available. It is therefore possible that a great wealth of information on a patient's allergy status may be entered electronically as free text in the form of symbols, words or abbreviations, but which cannot easily be transcribed into specific codes [11]. Such information is unwieldy, especially for use in whole population studies. The highest level of certainty is not always utilized by clinicians and this occurs particularly with asthma, where the majority of patients with asthma are found to be coded using, the Read code H33., a third tier concept with 16 available 'child' concepts of higher certainty available. When considering that this is typically likely to reflect the highest level of certainty required by a general practitioner to initiate treatment, this observation is perhaps not surprising. Indeed, it should be noted that although very detailed coded data entry in electronic medical records may be required for epidemiological research, it may well be adequate for clinical practice.

Table 2. Example of allergy codes used by different coding classifications

Coding systems	Allergy code (description)	Allergic asthma code (description)	Allergic rhinitis code (description)	Eczema/dermatitis code (description)
Clinical Terms Version 3	X79pp (allergy – mechanism)	X101x (allergic asthma)	XE0Y5 (allergic rhinitis)	X505K (allergic dermatitis)
International Classification of Primary Care-2	A12 (allergy/allergic reaction)	R96 (asthma)	R97 (hayfever/allergic rhinitis)	S87 (dermatitis/atopic eczema)
International Classification of Disease-10	T78.4 (allergy, unspecified) Z88 (personal history of allergy to drugs, medicaments and biological substances)	J45 (asthma) J46 (status asthmaticus)	J30 (allergic rhinitis)	L20 (atopic dermatitis) L23 (allergic contact dermatitis)
Read (4-byte version 2)	–	H431. (extrinsic asthma – atopy)	H28.. (hayfever) H29.. (allergic rhinitis NOS)	L23.. (atopic dermatitis/eczemas) L2z.. (dermatitis/eczemas NOS)
Read (5-byte version 2)	SN53. (allergy unspecified)	H330. (extrinsic atopic asthma) 1780. (aspirin induced asthma) H47y0 (detergent induced asthma)	H17.. (allergic rhinitis) H18.. (vasomotor rhinitis) Hyu20 (other seasonal rhinitis) Hyu21 (other allergic rhinitis) Hyu22 (other chronic sinusitis)	M11.. (atopic dermatitis) M12.. (contact dermatitis and other eczemas)
Systematized Nomenclature of Medicine-Clinical Terms	106190000 (allergic state)	389145006 (allergic asthma)	61582004 (allergic rhinitis)	24079001 (atopic dermatitis)

NOS, not otherwise specified.

Of particular concern is that ICD-10, ICPC and 5-byte Read (version 2) coding systems lack granularity and only provide a small number of concepts for diagnoses such as drug allergies [11, 12]. The hierarchical nature of ICD-10 and 5-byte Read codes has meant that it has been impractical to extend the range of codes to include all possible instances of drug allergy concepts, resulting in only 27 for ICD-10 and 57 for Read being included. Also, there may be some confusion as to the definition of adverse reactions to drugs in Read (Read code TJ. . .) with some clinicians using these codes for drug allergies. These problems may be resolved by linkage to a drug dictionary.

Further implications for reliance on these coding systems include difficulties in: producing reliable decision support systems whereby warnings may help to prevent inappropriate prescribing; populating referral documents and other communications, e.g. semi-automatic mail-merge extracts to facilitate reliable communication; and the onward transmission of patient information to other family practitioners or secondary/tertiary care.

Systematized Nomenclature of Medicine-Reference Terms (SNOMED-RT)

The College of American Pathologist's SNOMED-RT was derived in 1969 from two classification systems created

for pathology; it was designed to be a general purpose, comprehensive and computer-processable terminology [4]. SNOMED-RT has the capability of indexing all of the events found in a medical record (including prescribing) using alphanumeric terms. This is achieved through the use of 11 separate coding modules, which correspond to different axes of classification, such as diseases, procedures and morphology. For instance, a code exists for terms such as 'acute', 'new episode', 'child', 'skin or subcutaneous tissue structure' and 'atopic dermatitis' [3]. This system is no longer supported.

Clinical Terms Version 3

The Clinical Terms Version 3 (the Read codes) referred to as CTV3, was developed between 1992 and 1994 and aimed to provide a more comprehensive set of terms than 4-byte and 5-byte Read (version 2) [13]. This coding system encompasses specialist as well as general practice terms and is composed of several linked hierarchical structures (of unlimited depth), with some codes appearing several times. It has the ability to support the construction of more detailed concepts using a mechanism of qualifiers; for example, the core notion of 'allergic eczema' can be qualified by its exact site with reference to a detailed anatomy chapter [6]. It also has a clear structure

Table 3. International Classification of Disease-10 and 5-byte version 2 Read codes for allergic conditions

Clinical term	International Classification of Disease-10	5-byte version 2 Read codes
Allergic rhinitis	J30.1 (due to pollen)	H17.. and all codes below heading
Vasomotor rhinitis	–	H18..
Other seasonal allergic rhinitis	J30.2	Hyu20
Other allergic rhinitis	J30.3/4	Hyu21
Other chronic sinusitis	–	Hyu22
Allergic parotitis	–	J0720
Allergic gastritis	–	J1540
Allergic gastroenteritis and colitis	–	J432. and all codes below heading
Coeliac disease	–	J690. and all codes below heading
Allergic asthma	–	H330. and all codes below heading 1780. (aspirin induced asthma) H47y0 (detergent induced asthma)
All asthma (for comparison only)	J45 Asthma J46 Status asthmaticus	H33.. and all codes below heading 173A. (exercise induced asthma)
Wheezing	–	R0609 ([D] wheezing) 1737 (wheezing)
Urticaria	L50	M28.. and all codes below heading
Angioneurotic oedema	T78.3	SN51.
Hereditary angio-oedema	–	C3760
Anaphylaxis	–	SN50. and all codes below heading SP34. (due to serum)
Food allergy	L27.2 Dermatitis due to food T78.1 Other adverse reactions to food	SN58. and all codes below heading
Drug allergy	Z88 (personal history of allergy to drugs, medicaments and biological substances)	SN52. and all codes below heading ZV14 and all codes below heading 14L.. and all codes below heading
Venom allergy	T63.4	SN59. and all codes below heading
Allergy	T78.4 (allergy, unspecified)	SN53. and all codes below heading
Allergic conjunctivitis	H10.1	F4C14 (chronic) F4C06 (acute atopic) F4C14 (other chronic allergic conjunctivitis)
Contact or allergic eyelid dermatitis	–	F4D31
Chronic allergic otitis media	–	F5130
Allergic otitis media NOS	–	F5140
Allergic dermatitis and related	L23	M11..
Atopic dermatitis/eczema	L20	M111.
Infantile eczema	–	M112.
Flexural eczema	–	M113.
Allergic (intrinsic) eczema	–	M114. M11z. (atopic dermatitis NOS)
Neurodermatitis – atopic	–	M117.
Allergic contact dermatitis	–	M128. and all codes below heading
Dermatitis NOS	–	M12z0
Eczema NOS	–	M12z1
Infected eczema	–	M12z2
Hand eczema	–	M12z3
Erythrodermic eczema	–	M12z4
Allergic arthritis	–	N062. and all codes below heading
Anaphylactic shock	–	SN50. and all codes below heading
Skin: type 3 delayed reaction	–	3354.
Skin: type 1 immediate reaction	–	3355.
Skin: type 4 late reaction	–	3356.
Allergy skin test positive	–	3359.
Allergy test positive	–	3368.
History of food allergy	–	14M1.
Allergic purpura	–	D310. and all codes below heading
Allergic eosinophilia	–	D4033

NOS, not otherwise specified.

as it utilizes a defined syntax for permissible combinations [14]. This functionality provides much greater expressivity and allows specific codes to automatically inherit properties that exist in other hierarchies. For instance, the code pulmonary tuberculosis may be found by following both respiratory and infection hierarchies [3]. The concepts present in CTV3 have also been described as pure in that they are represented by unambiguous preferred terms and semantically correct synonyms with no duplication [13].

The SNOMED-CT

SNOMED-CT was formerly managed by the College of American Pathologists. However, responsibility for this coding system has now transferred to the new International Health Terminology Standards Development Organisation (IHTSDO), which is based in Denmark [15]. It is the intention of these participating countries that this system will become the primary compositional health care terminology to be used by all computers, this being driven by the recognition that the numeric coding structure of SNOMED-CT facilitates communications between health care professionals in clear and unambiguous terms and enables clinicians, researchers and patients from all countries intending to utilize this coding system to share and exchange health care and clinical information [16].

SNOMED-CT became available in January 2001, and is a combination of SNOMED-RT developed in the United States and CTV3 developed in the United Kingdom [6]. The coding system consists of several axes such as clinical finding, body structure and organism. These axes are hierarchical in structure and contain approximately 360 000 (35 000 clinical procedures) concepts. In addition, a mapping utility facilitates the translation of clinical concepts between different coding systems such as ICD-10 and several of the nursing terminologies to SNOMED-CT [7]. Another important feature of this system is the ability to compose complex post-coordinated concepts which represent a clinical meaning or code phrase built from a combination of several concepts. It is also possible to represent the same clinical meaning in several ways. Different standpoints and diseases, pathological processes, and anatomic sites, can be linked by qualifying expressions; for example, 'allergic contact dermatitis due to topical hydrocortisone allergy plus primary site: abdomen'. This system will thus allow detailed representation of clinical problems. It is suggested that the granularity provided by this system will have a fundamental long-term impact on clinical research [17].

SNOMED-CT has also been shown to be useful for decision support. An example of this is the TheraDocTM antibiotic assistant piloted within the secondary care environment by the University of Utah Health Science

Centre in Salt Lake City, USA [18]. This system utilized SNOMED-CT to translate clinical information including allergies, presenting symptoms, diagnoses and test results into a common language. This information was then used to create real-time patient summaries, patient safety and efficacy alerts, and individualized treatment recommendations (with precise drug and dosage information) that take into account considerations such as allergies, height, weight and renal function.

SNOMED-CT—potential problems with implementation

Although little formal assessment of SNOMED-CT has been undertaken, Elkin *et al.* [17] found that after trialling this coding system at the Mayo Clinic in Rochester, USA, a coding sensitivity of 92% was achieved, and that SNOMED-CT could represent a large proportion of terms seen commonly in medical problems lists. However, it was felt that further work was required to add missing synonyms. Similar work is currently underway to determine whether comparable sensitivity can be achieved in the United Kingdom, but in the meantime it is reasonable to conclude, based on the results of the Mayo Clinic study, that the compositional nature of SNOMED-CT has the potential of being able to include almost all allergy-related clinical terms. There is however a danger that nonsense relationships can also be built and that ambiguity and duplication could be introduced; all problems a controlled vocabulary intends to minimize. In addition, the introduction of SNOMED-CT with its flexibility and increased number of terms and concepts allows for clinicians to determine their own sets of codes useful to their individual clinical practice, but this is, in contrast, potentially very problematic for those wishing to carry out epidemiological interrogation of patient databases. In particular, this may significantly complicate the ability to undertake secondary analyses [14].

Other potential problems include the lack of a track record of trialling use of SNOMED-CT in routine clinical settings. It is, for example, generally recognized that common searches for terms should take less than a second to process; any greater time lag will cause user frustration. Also, although the IHTSDO intends to translate the SNOMED-CT coding system into different languages for international use, the direct linguistic translation of terms will need to be carried out with reference to culture-specific terminology.

The migration of other classification systems to SNOMED-CT will require significant work by users and their system suppliers. For instance, it is the intention of SNOMED-CT to 'retire' Read abbreviations such as NOS, which are commonly used [19]. It will be difficult to differentiate between concepts with similar names in small fields, particularly on hand-held devices where

users may find these fields difficult to read. Concerns have also been raised that implementation would raise major opportunity costs through the need for hardware and software, staff training (due to more detailed concepts and increased complexity) and resource impact. Also, some members of the primary and social care services have identified further gaps in the available concepts terminology, which would need to be addressed before release [20].

Recommendations

There are several issues that need to be resolved before the national implementation of SNOMED-CT (see Box 1). It is clear that agreeing terminology is work in progress. A process of normalization, whereby a computer system recognizes that two compositional constructs with the same meaning are the same so avoiding nonsense relationships, ambiguity and confusing codes needs to be carried out, and it is encouraging that this issue is currently being addressed by the IHTSDO [17].

Experts in allergy and coding should focus on setting up a working group to advise the newly formed IHTSDO, in order to examine potential gaps in the coding system for allergy diagnosis and management such as the lack of codes for drug and food allergy. Working groups have already been set up to tackle anaesthesia, anatomic pathology, dermatology, nursing, pharmacy and primary care. It is therefore important that professional bodies and government agencies within the countries that propose to adopt SNOMED-CT should encourage the setting up of such a group. This group should ensure that a defined, detailed and comprehensive set of allergy terms exists within SNOMED-CT that meet the needs of the allergy community [21]. Patients and expert bodies such as Scotland's National Clinic Dataset Development NHS

Programme, which support clinicians to develop interoperable national data standards, will also need to be involved before full implementation as failure to do so may lead to an inadequate allergy-related medical vocabulary once this system has been introduced. A standard set of codes, agreed across primary, secondary and tertiary care, will obviate some of the potential problems for those wishing to carry out secondary analysis. In addition, systems will need to be put in place so that secondary analyses are not jeopardised by the introduction of too many codes or terms. This standard terminology will need to integrate with all clinical systems, and this process will need to be managed centrally. The working group may also consider the benefits of incorporating the ICPC system, which incorporates codes for tertiary care and operates a clinically logical structure, bringing together symptoms and reasons for contact, diagnoses and interventions.

Sufficient resources and protected training time need to be made available to users, especially with regard to ensuring that structured coded data are consistently used. Where coding is being introduced as part of clinical practice for the first time (i.e. particularly in secondary and tertiary care), clinicians will need to be encouraged and trained in the use of SNOMED-CT before implementation.

It should also be possible to map data from external sources, such as social services, other coding systems and healthcare providers from other countries, onto SNOMED-CT. Mapping from these current systems will require that patient record searches retrieve all examples of target concepts, whether expressed in Read or the various ways possible in SNOMED-CT. So a standard equivalence detection mechanism is vital to aggregate reliable patient data and to provide automated decision support safely [22].

Box 1. Solutions to the potential problems posed by implementation of SNOMED-CT

Potential problems	Potential solutions
Gaps in coding	Experts in allergy and coding should focus on setting up a working group to advise the newly formed International Healthcare Terminology Standards Development Organisation to examine potential gaps in the coding system
Mapping from current systems	Mapping from current systems will require that patient record searches retrieve all examples of target concepts, whether it is expressed in Read or the various ways possible in SNOMED-CT
Complex concepts	Wherever possible, new or post-coordinated concepts and codes should be determined by allergy experts working with government bodies. Codes need to be trialled by simulation. The commonest used codes should be identified and highlighted for inclusion
Training	Sufficient resources and protected training time should be made available to users, especially with regard to ensuring that structured and coded data are used, and that codes and terms are used consistently. Training will need to be prioritised to these groups new to coding as part of clinical practice
Secondary analyses	A standard set of codes will obviate some of the problems for those wishing to carry out secondary analyses. In addition, systems will need to be put in place so that secondary analyses are not jeopardised by the introduction of too many new codes or terms

SNOMED-CT, Systematized Nomenclature of Medicine-Clinical Terms.

Conclusions

The analysis of data collected through the use of high quality computerized systems is important to better understand and continue to monitor the health burden posed by allergic diseases. Current coding systems have been found to be clinically restrictive and inadequate for developing a comprehensive picture of the true disease burden posed by allergic conditions. Greater granularity of coding can be achieved by using the SNOMED-CT system and this system is therefore likely to be welcomed by clinicians. However, it is important that, before the introduction of SNOMED-CT into healthcare systems, several issues are resolved, these including: a need to ensure that adequate mapping occurs from existing systems; that SNOMED-CT is trialled sufficiently before implementation; and that training is provided to users new to coding as part of their clinical practice. It is vital that an allergy working group, drawing on the expertise of clinicians and representatives of the IHTSDO, is formed, with a remit to bring to light any gaps in allergy coding and work towards developing a set of preferred codes that can be recommended for routine use so as to allow high quality secondary analyses of routinely collected data to continue.

Acknowledgements

Funding: Chief Scientist's Office of the Scottish Executive. Ref: CZG/2/252.

Potential conflict of interests: A. S. was part of the Department of Health Allergy Review Stakeholder Group, has given evidence to the House of Lords Inquiry, and together with CF serves on the Scottish Executive's Review of Allergy Services in Scotland Working Group. A. S. is a member of the Programme Advisory Group for the Connecting for Health Evaluation Programme.

References

- Gupta R, Sheikh A, Strachan DP, Anderson HR. Burden of allergic disease in the UK: secondary analyses of national databases. *Clin Exp Allergy* 2004; 34:520–6.
- Anandan C, Simpson CR, Fischbacher C, Sheikh A. Exploiting the potential of routine data to better understand the disease burden posed by allergic disorders. *Clin Exp Allergy* 2006; 36:866–971.
- Coiera E. *Guide to medical informatics, the internet and telemedicine*. London: Chapman & Hall, 1997.
- Roscoe T. Classification systems: ICD, SNOMED, Read codes. Available from: <http://www.wisdom.org.uk/class.html> (last accessed 14 November 2006).
- Donaldson RJ, Donaldson LJ. *Essential Public Health Medicine*. Lancaster: Kluwer Academic Publishers, 1993.
- De Lusignan S, Van Weel C. The use of routinely collected computer data for research in primary care: opportunities and challenges. *Fam Pract* 2006; 23:253–63.
- Coonan KM. Medical informatics standards applicable to emergency department information systems: making sense of the jumble. *Acad Emerg Med* 2004; 11:1198–205.
- Graham L, Ward A, Mulvenna G. *User's guide to Read codes*. Edinburgh: Her Majesty's Stationary Office, 2000.
- Gray J, Orr D, Majeed A. Use of Read codes in diabetes management in a south London primary care group: implications for establishing disease registers. *BMJ* 2003; 326: 1130–34.
- NHS National Services Scotland. National Clinic Dataset Development Programme. Available from: <http://www.clinicaldatasets.scot.nhs.uk> (last accessed 2 October 2006).
- Horsfield P. Clinical advisor lead, connecting for health. Personal communication, 29 September, 2005.
- Scottish Public Health Observatory. Allergic conditions: ICD10 codes used for analysis in secondary care. Available from: http://www.scotpho.org.uk/web/site/home/Healthwell-beinganddisease/Allergic_conditions/Allergicconditionsdata/Allergic_conditions_ICD10codes.asp (last accessed 14 September 2006).
- Brown PJB, Warmington V, Laurence M, Prevost AT. Randomised crossover trial comparing the performance of Clinical Terms Version 3 and Read Codes 5 byte set coding schemes in general practice. *BMJ* 2003; 326:1127–31.
- Campbell JR, Carpenter P, Sneiderman C, Cohn S, Chute CG, Warren J. Phase II evaluation of clinical coding schemes: completeness, taxonomy, mapping, definitions, and clarity. *J Am Med Inform Assoc* 1997; 4:238–51.
- SNOMED International, a Division of the College of American Pathologists. Available from: www.snomed.org (last accessed 10 October 2007).
- The National Health Service. Connecting for health. Available from: <http://www.connectingforhealth.nhs.uk/terminology/snomed> (last accessed 16 November 2006).
- Elkin PL, Brown SH, Husser CS *et al*. Evaluation of the content coverage of SNOMED CT: ability of SNOMED Clinical Terms to represent clinical problems problem lists. *Mayo Clin Proc* 2006; 81:741–8.
- SNOMED Clinical Terms (SNOMED CT) in TheraDoc's Antibiotic Assistant underlines actionable clinical decision support. Available from: www.snomed.org (last accessed 19 December 2006).
- Caunce D, Crook D, Harding J, Green I. *Convergence of OPCS-4 and SNOMED-CT: feasibility study – stage 1 report*. London: Crown Copyright, 2006.
- Ramsay L, Ward A. *SNOMED clinical terms: Scottish evaluation report*. Edinburgh: NHS Scotland, 2003.
- Chalmers RJG. Health care terminology for the electronic era. *Mayo Clin Proc* 2006; 81:619–24.
- Herbert I, Hawking M. Bringing SNOMED-CT into use within primary care: report of the meeting of the clinical computing special interest group (CLICSIG) of the primary health care specialist group of the British computing society. *Inform Prim Care* 2004; 13:61–4.