



[HIMJ HOME](#)

[CONTENTS](#)

[GUIDELINES](#)

[MISSION](#)

[CONTACT US](#)

HIMAA  
Locked Bag 2045  
North Ryde, NSW  
Australia 1670  
ABN 54 008 451 910

## HIMJ: Reviewed articles

### Matching GP terms to the ICD-10-AM index

Peter R Scott

#### Abstract

This article describes some preliminary work done by the National Centre for Classification in Health (NCCCH) in response to the recommendations of the General Practice Coding Jury. Terms derived from two Australian general practices, two coordinated care trials and the General Practice Coding Jury were matched to the index of the *International Statistical Classification of Diseases and Related Health Problems, 10th Revision, Australian Modification* (ICD-10-AM) to gauge the content coverage provided by the existing index. Results showed the percentage of good general practice (GP) term/ICD-10-AM Index term matches was 58%. The percentage of acceptable concept matches was 88%. It is concluded that this work provides a useful methodology, and that the ICD-10-AM index may support a level of concept matches, while it will require augmentation to support term matching.

**Key Words:** Australia; classification; family practice; vocabulary, controlled; coding system

#### Introduction

##### Aim

The aim of the study was to assess the breadth or comprehensiveness of the ICD-10-AM index for terms used in Australian general practice. This was to help NCCCH develop both an understanding of the content coverage of ICD-10-AM for the Australian general practice domain, and a methodology to be incorporated into the Australian alpha trials of the Systematized Nomenclature of Medicine – Clinical Terms (SNOMED-CT).

##### Terms and concepts

In this article, 'term' refers to the diagnostic word or phrase used in the clinical documentation. The word 'concept' means that to which the term refers. It is possible to have more than one term for each concept and it is accepted that this occurs in a significant way in general practice. An example is the use of the terms 'runny nose' and 'rhinorrhoea' for the same concept.

##### Background and literature review

In 2000, the peak body in Australia for general practice informatics recommended an appropriate coding system for Australian general practice ([General Practice Coding Jury 2000](#)). Specifically, the General Practice Coding Jury proposed that Australia adopt an augmented ICD-10-AM clinical coding scheme for the next five years, and become involved in the development of SNOMED-CT.

NCCH has incorporated the tabular list of ICD-10-AM diseases and procedures into the Unified Medical Language System (UMLS), commencing in 1999 ([Scott 2001](#)). One of the outcomes of this work was the realisation that the alphabetical index also needed to be included, as it contains synonyms and Australian terms that represent a breadth of content coverage not encompassed by the codes or the code titles to which they are referenced. It was anticipated that the index would be most suitable for general practice, as practitioners in this domain tend to use a wide variety of terms idiosyncratically ([Crombie 1992](#)).

The work presented in this article furthers the recommendations of the General Practice Coding Jury as well as providing an evaluation of NCCH's hypothesis, specifically that the ICD-10-AM index is comprehensive for Australian general practice terms.

### Methods

A GP termset was obtained from various sources. Firstly, the 50 most frequent terms from a six-doctor practice and a four-doctor sister clinic<sup>[1]</sup> were retrieved from the practices' clinical software. This software is 'Medical Director'.<sup>[2]</sup> Secondly, after permission was obtained from the data custodians of the Brisbane North TEAMCare Health in the state of Queensland and the Illawarra Coordinated Care trials in the state of New South Wales ([Commonwealth Department of Health and Aged Care 1999](#)), de-identified diagnostic terms were sampled. The Coordinated Care data sources had different media for data collection. The Brisbane North-based TEAMCare Health trial had paper sources containing original patient enrolment forms from participating general practitioners. Three hundred records were read and up to 10 diagnostic terms and phrases per client were dictated onto tape. These data were then transferred into an Excel spreadsheet. This trial had an average of five enrolments per general practitioner.

The Illawarra data were in electronic format. The data for this trial were collected via a modified 'Medical Director' package used by participating general practitioners in that trial. The number of general practitioners contributing to the data set is not known. Finally, the terms from the General Practice Coding Jury assessment process were also included. [Box 1](#) shows information about the data sources. Factors considered in assessing the quality of the data were age, sample size, internal validity and likelihood of being representative of the domain of interest. The ideal data source would have been recent, with a representative sample size and 'clean' data containing no apparent inconsistencies or errors, and the content would be consistent with the terms generated by general practitioners in their work.

#### 1: Source metadata

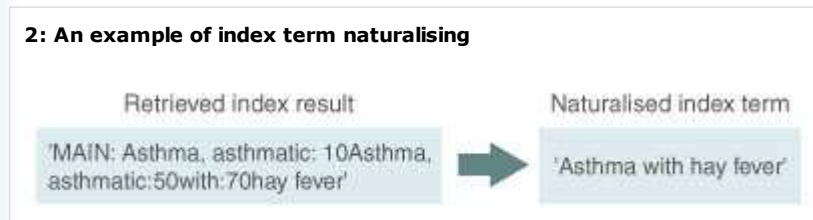
Source	Years data collated	Quality of data	Number of terms	Sampling

Practice A	2000–2001	Moderate	50	Top 50 from 1054 terms
Practice B	2000–2001	Moderate	50	Top 50 from 632 terms
TEAMCare Health Coordinated Care Trial	1998–1999	Good	210	All terms used more than once from 816 terms
Illawarra Coordinated Care Trial	1998–1999	Moderate	150	All terms used more than four times from a set of 1613 terms
Coding Jury sample data	Not known	Good	66	66

Most of the sources have used Medical Director as their clinical software. This software incorporates the Docle classification system ([Oon 1988](#)). In this environment, the terms are usually added at the time of prescribing, although terms also may be added in medical history fields.

Coding of the terms, using ICD-10-AM, was undertaken by professional coders employed by NCCH. The assigned codes often constituted a range of codes if the GP term was a general term. A Structured Query Language (SQL) query on the NCCH ICD-10-AM database retrieved all index terms incorporating the chosen codes or code ranges. The author then naturalised the appropriate index terms for the GP term and chose a preferred term from the index terms. 'Naturalised' in this sense means that a natural language equivalent was chosen for the retrieved index term. This process of naturalising was done to enhance the capacity of the index to become useful as an interface terminology. This would allow a future user of an electronic index to identify their preferred terms and use them as an alternative 'way-in' to the classification. The original index would be preserved. It is important to note different users may prefer different views of the index. An example is presented in Box 2.

### 2: An example of index term naturalising



A rating was assigned for the quality of the GP term–Index term match. These ratings are detailed in Box 3. Chi square tests were used to assess the significance of the difference in ratings for each data source.

### 3: Ratings and what they mean

Rating	Meaning
1	Matched GP term-10AM index term and code assigned
2	GP term-10AM index term do not match but relevant code available
3	No 10-AM index term or relevant code

Data from the various sources were entered into a spreadsheet with the following column headings: source, GP term, ICD-10-AM assigned code(s), rating of match, count of index terms, frequency of term per source, and comments.

### Results

The total number of unique terms for all sources was 374. Box 4 outlines the top 10 terms per source. The data from the general practices were combined, as there was a high level of commonly used terms in both. Acronyms were preserved, as these are valuable GP terms in their own right.

#### 4: Top 10 terms per source

<u>Combined practices</u>	<u>TEAMCare Coordinated Care trial</u>	<u>Illawarra Coordinated Care trial</u>
URTI	Hypertension	Fluvax
	Osteoporosis	Hypertension
Hypertension		
	Ischaemic heart disease	Chest infection
Asthma		
	Osteoarthritis	BP check
Osteoarthritis		
	Depression	Asthma
LRTI		
	NIDDM	Insomnia
Depression		
	Asthma	Back pain
Sinusitis		
	Anxiety	Sciatica
Hyperlipidaemia		
	Atrial fibrillation	Dog bite
Insomnia		
	Hypothyroidism	Solar keratosis cryo
GORD		

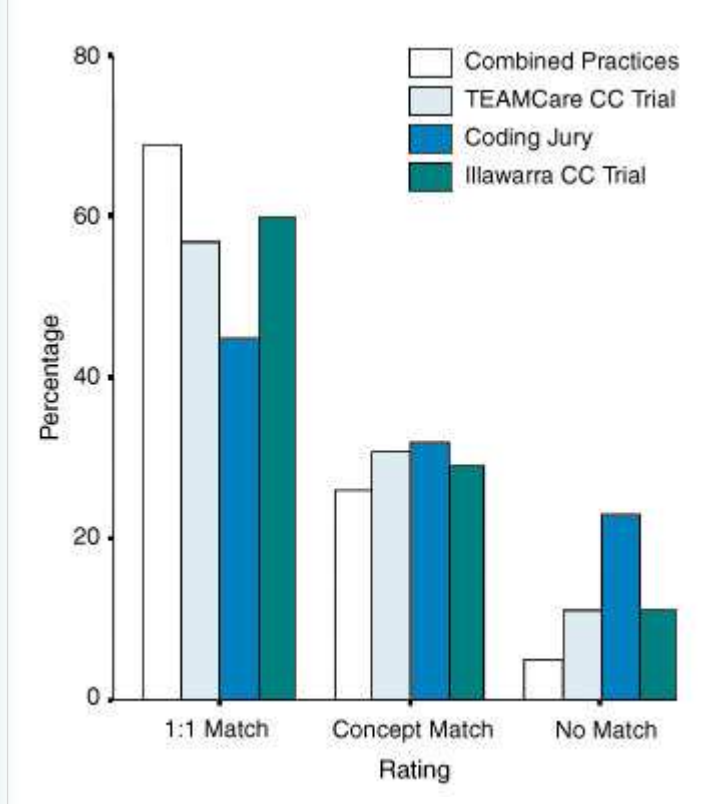
Box 5 details the numbers of terms in each trial and the breakdown of the ratings scores, and presents these data as percentages of the total number of terms per source. On average, rating 1 term matches occurred in 58% of cases, rating 2s occurred in 30% of cases, and rating 3s occurred in 12% of cases.

**5: Rating scores and percentages for conjoint source subsets**

Source	Total terms	Rating 1s (%)	Rating 2s (%)	Rating 3s (%)
Combined practices	62	43 (69)	16 (26)	3 (5)
TEAMCare CC trial	210	120 (57)	66 (31)	24 (11)
Coding Jury	66	30 (45)	21 (32)	15 (23)
Illawarra CC trial	150	90 (60)	43 (29)	17 (11)
Average %	-	- (58)	- (30)	- (12)

Box 6, which presents similar information to that shown in Box 5, demonstrates the percentage of ratings across the four data sources.

**6: Rating scores and percentages for conjoint source subsets**



Chi square tests were used to assess the significance of the differences in ratings for each data source (Box 7 and Box 8). Box 7 presents differences between 1:1 matches (rating 1) and no matches. Box 8 presents both 1:1 and 1:n matches, compared with no matches.

In Box 7, there is a significant difference for the combined practices between the frequency of ratings, with 1:1 matches (rating 1s) occurring significantly more frequently than no matches (rating 3s) ( $\chi^2 = 40.29, p < 0.05$ ). For TEAMCare, there is a significant difference between the frequency of ratings, with 1:1 matches (rating 1s) occurring significantly more frequently than no matches (rating 3s) ( $\chi^2 = 66.17,$

$p < 0.05$ ). For Coding Jury data, there is no significant difference between the ratings frequency ( $\chi^2 = 5.182, p = >0.05$ ). Therefore, equally as many terms were rated as matched (rating 1s), concept matched (rating 2s), and not matched (rating 3s). For the Illawarra data, there is a significant difference between the frequency of ratings, with 1:1 matches (rating 1s) occurring significantly more frequently than not relevant codes (rating 3s) ( $\chi^2 = 54.76, p < 0.05$ ).

**7: Statistical results comparing 1:1 matches and no matches for each data source**

<u>Data source</u>	<u>N</u>	<u><math>\chi^2</math></u>	<u>df</u>	<u>p</u>
Combined practices	62	40.290	2	0.001*
TEAMCare CC trial	210	66.170	2	0.001*
Coding Jury	66	5.182	2	0.075*
Illawarra CC trial	150	54.760	2	0.001*

\* $p < 0.05$

In Box 8, there is a significant difference for the combined practices between the frequency of ratings, with 1:1 matches (rating 1s) together with 1:n matches (rating 2s) occurring significantly more frequently than no matches (rating 3s) ( $\chi^2 = 50.58, p < 0.05$ ). For TEAMCare, there is a significant difference between the frequency of ratings, with 1:1 matches (rating 1s) together with 1:n matches (rating 2s) occurring significantly more frequently than no matches (rating 3s) ( $\chi^2 = 124.97, p < 0.05$ ). For Coding Jury data, there is a significant difference between the ratings, with 1:1 matches (rating 1s) together with 1:n matches (rating 2s) occurring significantly more frequently than no matches (rating 3s) ( $\chi^2 = 19.63, p < 0.05$ ). For the Illawarra data, there is a significant difference between the frequency of the ratings, with 1:1 matches (rating 1s) together with 1:n matches (rating 2s) occurring significantly more frequently than no matches (rating 3s) ( $\chi^2 = 89.70, p < 0.05$ ).

**8: Statistical results comparing 1:1 and 1:n matches with no matches for each data source**

<u>Data source</u>	<u>N</u>	<u><math>\chi^2</math></u>	<u>df</u>	<u>p</u>
Combined practices	62	50.58	1	0.001*
TEAMCare CC trial	210	124.97	1	0.001*
Coding Jury	66	19.63	1	0.001*
Illawarra CC trial	150	89.70	1	0.001*

\* $p < 0.05$

Box 9 shows the average and median counts of the index

terms per GP term. For each GP term there were between 0 and a maximum of 539 retrieved index terms. It was important to measure this, as NCCH needed to see if the methodology could be sustained. That is, the amount of work involved in having a clinician review each retrieved index term for a given GP term needed to be measured in some way.

**9: Median and average counts of ICD-10-AM index terms for each GP term**

	<u>N</u>	<u>Median count</u>	<u>Average count</u>
All unique terms	374	11	30
1s	202	13	33
2s	117	13	31
3s	55	6	17
1s and 2s	319	13	32
2s and 3s	172	10	27

Box 10 shows the median and average counts by source. The relevance of these data pertains to whether terms sourced from general practice retrieve more or less ICD-10-AM index terms than those sourced from an exercise such as a coordinated care trial enrolment. In the latter, the cohort of patients may possess less heterogeneous disease states than those of the entire community.

**10: Median and average counts, by source**

<u>Source</u>	<u>Median count</u>	<u>Average count</u>
Combined Practices	31	55
TEAMCare Coordinated Care trial	15	37
Illawarra Coordinated Care trial	13	33

### Discussion

Overall, the results show that more than half the GP terms had a term match with the index, and a further 30% had a concept match. This suggests that the index is capable of augmentation for a more general and representative GP termset.

An average of 30 index term entries retrieved per GP term is of note. This supports the notion that the ICD-10-AM index contains a breadth of terms suitable for general practice. Additional evidence is in the larger number of retrieved index terms per GP term for data sourced from a reason-for-prescribing prompt (ie, Medical Director) than for a more formal enrolment process for a specific patient cohort (eg, a coordinated care trial).

The sample obtained was not suitable for the task of establishing a comprehensive GP termset. A preferred national GP Vocabulary is likely to contain a number of terms much larger than that obtained. In addition to this, the sample obtained is biased towards the frail aged domain. The nature of the study, however, was preparatory and was helpful in allowing NCCH to examine a particular methodology.

Some caution is warranted regarding the frequency data in the Illawarra Coordinated Care trial data set supplied. The integrity of the data is questionable, as the top 10 terms include the term 'dog bite'. Further research would be indicated to determine if this is indeed the case.

Another potential source of error may be a function of having ratings determined by the author alone, allowing for potential bias. Future research may use a team of classification experts for this task, where inter-coder reliability may be assessed and measures may be employed to avoid scoring drift, which has weakened other studies ([Chute et al 1996](#)).

### Conclusions

In conclusion, the ICD-10-AM index appears to support concept matches and an augmented index is likely to support lexical GP-term matches with a larger, more representative GP termset. Augmentation is feasible and the preliminary methodology established appears to be robust.

### Acknowledgments

The author acknowledges the help of Kirsten McKenzie of NCCH Brisbane with the statistics and with helpful comments, and Donna Truran and Michelle Bramley, of NCCH Sydney, and Sue Walker, of NCCH Brisbane, for their helpful comments. Thanks also go to Kerry Innes, Sheree Gray and Jenny Seems, of NCCH Sydney, for their help with coding.

### References

Chute CG, Cohn SP, Campbell KE and Campbell JR (1996). 'The content coverage of clinical classifications for the computer-based Patient Record Institute's work group on codes & structures'. *Journal of the American Medical Informatics Association*, 3(3):224-233.

Commonwealth Department of Health and Aged Care (1999). *The Australian Coordinated Care Trials: Background and Trial Descriptions*. Commonwealth Canberra: The Department.

Crombie DL, Cross KW and Fleming DM (1992). 'The problem of diagnostic variability in general practice'. *Journal of Epidemiology and Community Health*. 46(4): 447-454.

General Practice Coding Jury (2000). *Final report of the General Practice Coding Jury*. Canberra: Commonwealth Department of Health and Aged Care.

Oon YK (1988). 'DOCLE. A new system of medical notation'. *Australian Family Physician* 17(9): 772-773.

Scott PR (2001). 'Incorporation of ICD-10-AM into the UMLS'. *Proceedings of the 11th RACGP Computer Conference*.

### Readers may be interested in the following papers that have been presented by Dr. Scott:

Scott PR and Schulz E (1998). 'Four possible coding systems for Australian general practice: A formal evaluation'. *Proceedings of the Health Informatics Conference (HIC1998)*.

Scott PR (2001). 'Incorporation of ICD-10-AM into the UMLS'. *Proceedings of the 11th RACGP Computer Conference*.

**Peter R Scott MBBS, BA**

Project Officer, National Centre for Classification in Health (Brisbane),  
School of Public Health, Queensland University of Technology.  
General Practitioner, Brisbane, Queensland.  
Tel: 07-3864-5809 E-mail: [pr.scott@qut.edu.au](mailto:pr.scott@qut.edu.au)

NCCH (Brisbane) is funded by the Australian Bureau of Statistics, the Australian Institute of Health and Welfare, Queensland University of Technology, and the Commonwealth Department of Health and Aged Care through NCCH (Sydney).

---

[1] Three of the doctors worked in both locations.

[2] <[www.australiandoctor.com.au/md.asp](http://www.australiandoctor.com.au/md.asp)>



DO YOU KNOW WHAT'S GOING ON?

---

© 2002 Health Information Management Association of Australia Limited