



[HIMJ HOME](#)

[CONTENTS](#)

[GUIDELINES](#)

[MISSION](#)

[CONTACT US](#)

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

HIMJ Reviewed articles

Assessment of the impact of the change from manual to automated coding on mortality statistics in Australia

Kirsten McKenzie, Sue Walker and Shilu Tong

Abstract

It remains unclear whether the change from a manual to an automated coding system (ACS) for deaths has significantly affected the consistency of Australian mortality data. The underlying causes of 34,000 deaths registered in 1997 in Australia were dual coded, in ICD-9 manually, and by using an automated computer coding program. The diseases most affected by the change from manual to ACS were senile/presenile dementia, and pneumonia. The most common disease to which a manually assigned underlying cause of senile dementia was coded with ACS was unspecified psychoses (37.2%). Only 12.5% of codes assigned by ACS as senile dementia were coded the same by manual coders. This study indicates some important differences in mortality rates when comparing mortality data that have been coded manually with those coded using an automated computer coding program. These differences may be related to both the different interpretation of ICD coding rules between manual and automated coding, and different co-morbidities or co-existing conditions among demographic groups.

Keywords: classification; coding practice; data quality; mortality

Introduction

Mortality data play an important role in informing public health policy, setting health priorities and allocating resources, and assessing health services quality and outcomes (Taylor et al, 1998; Johansson & Ragnar, 2000). Many factors can affect the consistency of mortality data over time, including those relating to the collection, classification and processing of the data. The introduction of an Automated Coding System (ACS) for deaths registered from 1 January 1997 was one of the major changes in Australian mortality data processing occurring in recent years (ABS, 2000).

Before the introduction of ACS, coding was undertaken manually, with coders from the Australian Bureau of Statistics being responsible for assigning International Classification of Diseases (ICD) codes to represent the underlying cause of death from the death certificates completed by clinicians and coroners. Several problems were associated with this manual approach, with the major difficulties being variability between coders in the interpretation of causal sequences reported on death certificates, inconsistent assignment of underlying cause codes, and the inability to code and report on multiple causes of death (Israel, 1990).

The availability of multiple cause of death data is becoming increasingly important for public health research and development. Israel (1990)

suggests that as deaths from chronic diseases become more significant than deaths from infectious or parasitic diseases, it becomes more important to capture data about multiple causes of death, as coexisting conditions may contribute significantly to a death. In addition, as the world population ages, deaths are due increasingly to a complex mix of conditions, none of which may be uniquely identifiable, necessarily, as the underlying cause of death. In the USA, while only 20% of deaths at the beginning of the 1900s were due to chronic diseases such as heart disease and cancer, approximately 70% of all deaths in 1990 were due to these diseases. Greig (1999) reported that prior to the use of ACS in Australia, an average of 1.8 causes, per death, were lost using manual coding and single underlying cause-of-death data.

The ACS was developed originally in the late 1960s by the United States National Center for Health Statistics. The aim of developing ACS was to meet the need for reporting multiple causes of death and to aid in standardising the interpretation and coding of causes of death (Israel, 1990). The ACS has several advantages compared with manual coding. First, ACS makes it relatively easier to determine the underlying cause of death. Usually, data from the death certificate are entered into the computer in standard format and the ACS then uses a series of decision tables (based on the current ICD rules) to identify which of the diseases on the death certificate is most likely to be the underlying cause of death. Second, other conditions (*ie.* multiple causes of death), also can be identified and coded so that a far richer data set can be produced. Third, while individual coders may vary in their interpretation of the ICD rules, the automated coding software provides a consistent method for coding cause of death data. Finally, as all decisions applied by the ACS are clearly reported, troubleshooting to solve technical problems is quite straightforward (Israel, 1990).

The ACS has been adopted by a number of countries, including Australia, because it has apparent benefits as mentioned above. Its growing popularity has the potential to improve significantly the international comparability of causes of death statistics (ABS, 2000). However, although the inherent benefits of ACS have been widely acknowledged, the impacts on mortality data of the change from manual to ACS have been noticed. For example, one of the most obvious effects has been differences in the interpretation of ICD cause-of-death coding rules when comparing the output from manual and automated coding (Greig, 1999). Prior to ACS, manual coders used an 'Australian' interpretation of the ICD rules, which tended to be more restrictive than that used in the ACS. In order to overcome this problem, the Australian Bureau of Statistics has calculated comparability factors for groups of causes of death as a means of adjusting data for 1997 and later years (ABS, 2000). However, a comprehensive assessment of the potential impact of such change on the mortality data has not yet been reported in the literature.

This study aims to examine the major discrepancies in death coding between manual and ACS, and their variations with demographic factors.

Method

We obtained a data file containing approximately 34,000 deaths registered in 1997 that were dual coded by the Australian Bureau of Statistics, manually and in ACS, using ICD-9. These deaths represent approximately a quarter of all deaths for that year. Comparability factors were produced for diseases by dividing the number of deaths

coded to a certain underlying cause under ACS by the number of deaths due to that same cause under the manual system (ABS, 2000). Comparability factors of one or close to one indicated little change in the number of deaths coded to a particular cause between the manual and automated systems, which represented a good comparability between two systems. Comparability factors of less than one indicated that there were fewer deaths coded to a certain disease under the automated system than the manual system, while comparability factors of greater than one indicated that there were more deaths coded to a certain disease under the automated system than the manual system.

The diseases with either the lowest or highest comparability factor were examined in more detail to ascertain which codes were assigned by the automated coding system compared with the manually assigned codes. We also examined the differences in comparability factors across different demographic groups (*ie.* age and sex) for the diseases of interest. Comparability factors and standard errors for different groups were calculated. Standard errors of proportions were produced for each specific age group and by sex using the formula $\sqrt{p[1-p]}/n$ (where p = number of deaths coded the same using ACS/total number of manual deaths for certain disease).

Results

[Table 1](#) shows the number of cases and percentage of data assigned the same codes by both manual and automated processes for diseases at the chapter level. Good concordance between manual and automated coding occurred for pregnancy/childbirth (100%), neoplasms (98.7%), congenital diseases (96.0%), and circulatory system disorders (95.6%). Poorer concordance was found for the following diseases: symptoms/signs (73.2%), mental disorders (74.8%), blood diseases (82.2%), genito-urinary system (83.4%), and musculoskeletal system (85.3%).

Figure 1: Percentage of data coded the same from manual to automated coding

ICD Chapter Level Diseases	N	Same Code	Same Code (%)
Infectious/Parasitic	384	362	94.3%
Neoplasms	8849	8737	98.7%
Blood	101	83	82.2%
Endocrine/Nutritional	1138	1031	90.6%
Mental/Behavioural	1040	778	74.8%
Nervous System	819	702	85.7%
Circulatory System	14817	14164	95.6%
Respiratory System	3408	3227	94.7%
Digestive System	1028	959	93.3%
Skin/Subcutaneous	68	63	92.6%
Musculoskeletal System	184	157	85.3%
Genitourinary System	631	526	83.4%
Pregnancy/Childbirth	5	5	100%
Perinatal Diseases	142	127	89.4%

Congenital Diseases	173	166	96.0%
Symptoms/Signs	168	123	73.2%

(Note: As the majority of external causes were manually coded rather than using ACS (because ACS is problematic when used for external causes), the 'External Cause' chapter and the 'Injuries and Poisoning' chapter are not included in this table).

[Back to text](#)

The manual to automated differences are shown in [Figure 1](#). There is a remarkable discrepancy between manual and automated coding systems for some diseases. For example, 83.8% of the cases manually coded as senile/presenile organic psychotic conditions (manual code: 290) were coded differently by the automated coding system. However, 57.9% of the cases coded as senile/presenile organic psychotic conditions by ACS were coded differently by coders.

Figure 1: Percentage of codes different between manual and automated coding systems

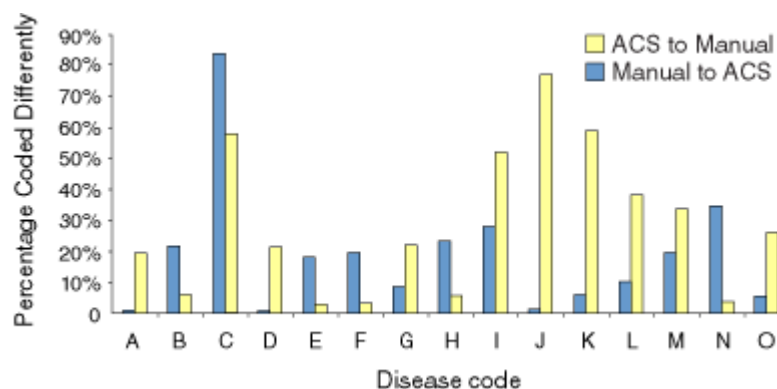


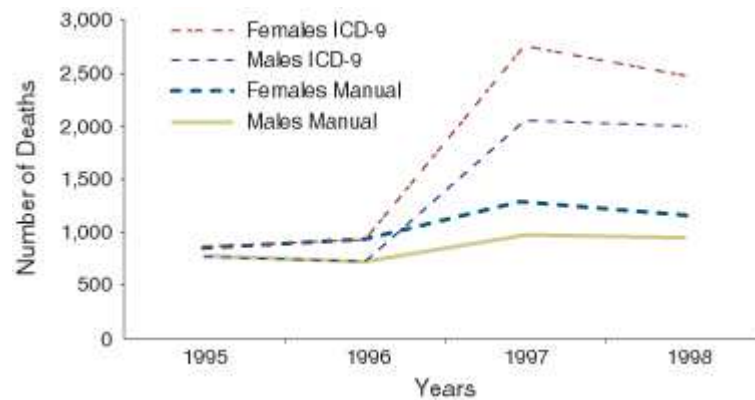
Figure 1 code references

- | | |
|---|--|
| A = Malignant neoplasm of connective/soft tiss (171) | I = Pneumococcal pneumonia(481) |
| B = Neop of uncertain behav other unspec sites (238) | J = Bronchopneumonia (485) |
| C = Senile and presenile organic psychotic (290) | K = Pneumonia (486) |
| D = Alcohol dependence syndrome (303) | L = Nephritis and nephropathy (583) |
| E = Other cerebral degeneration (331) | M = Acute renal failure (584) |
| F = Parkinson's disease (332) | N = Chronic renal failure (585) |
| G = Chronic pulmonary heart disease (416) | O = Other congenital anomalies of CNS (742) |
| H = Occlusion of cerebral arteries (434) | |

[Back to text](#)

[Figure 2](#) shows that the mortality rates for pneumonia for males and females would have differed in 1997 if ACS were not introduced. A large discrepancy in the data is evident at the point when ACS was introduced.

Figure 2: Pneumonia mortality rates with and without ACS



[Back to text](#)

[Figures 3](#) and [4](#) show the major coding changes between manual and automated coding for senile dementia (ICD Code 2900) and bronchopneumonia (ICD Code 4859), respectively. For example, of the total 726 senile dementia cases originally manually coded, only 94 (12.9%) were coded as senile dementia by the ACS. Two hundred and seventy cases (37.2%) were coded by ACS as unspecified psychoses, 92 (12.7%) as pneumonia, and 67 (9.2%) as bronchopneumonia ([Figure 3](#)). Similarly, of the 534 bronchopneumonia cases coded by the ACS, only 212 (39.7%) were coded the same manually. Sixty-seven (12.5%) were manually coded as senile dementia, and 22 (4.1%) as ischaemic heart disease ([Figure 4](#)). Sex-specific differences in the coding between the two systems are also shown in these figures.

Figure 3: Manual code senile dementia (2900) compared to automated codes

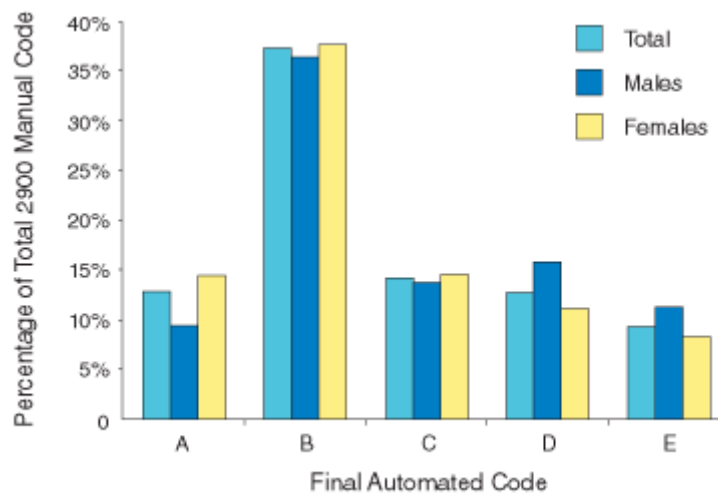


Figure 3 code references

- | | |
|----------------------------------|-----------------------------|
| A = Senile dementia | D = Pneumonia |
| B = Unspecified psychoses | E = Bronchopneumonia |
| C = Presenile dementia | |

[Back to text](#)

Figure 4: Final auto code bronchopneumonia (4859) compared to manual codes

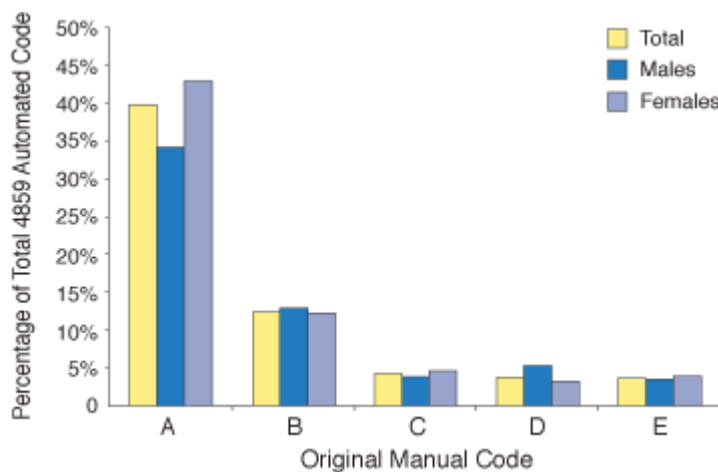


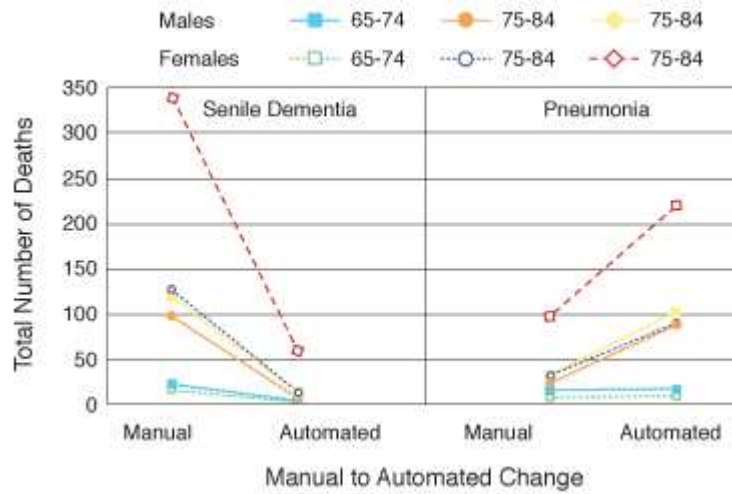
Figure 4 code references

- | | |
|-----------------------------|---------------------------------|
| A = Bronchopneumonia | D = Hereditary/degen CNS |
| B = Senile dementia | E = Heart failure |
| C = Ischaemic HD | |

[Back to text](#)

[Figure 5](#) shows the number of deaths using manual and automated coding by disease and demographic group. It appears that ACS coded fewer deaths from senile dementia but more from pneumonia than manual coding, particularly in the female elderly (*ie.* over 85 years).

Figure 5: Number of deaths using manual and automated coding by disease and demographic



[Back to text](#)

We also examined comparability factors and standard errors for senile/presenile dementia and bronchopneumonia by age and sex. [Table 2](#) shows that the overall comparability factor is 0.38 for senile/presenile dementia, and 2.48 for bronchopneumonia. Comparability factors varied with age and sex.

Table 2: Comparability factors (standard errors) by age and sex

	Males			Females			Total
	65-	75-	85-	65-	75-	85-	65-
Senile/Presenile Dementia [N = 767]	0.48	0.24	0.26	0.45	0.26	0.34	0.38
	(0.09)	(0.04)	(0.04)	(0.11)	(0.04)	(0.02)	(0.01)
Bronchopneumonia [N = 215]	1.00	4.35	3.25	1.14	2.96	2.30	2.48
	(0.00)	(0.41)	(0.27)	(0.14)	(0.25)	(0.11)	(0.01)

[Back to text](#)

Discussion

The change from manual to automated coding has presented a number of challenges to the Australian system for coding mortality data, although the introduction of ACS allows a more consistent approach to the assignment of cause of death codes, and the ability to code multiple causes of death information in a functional format. It has been

recognised that there is scope for differences between manual and automated coding systems (ABS, 1997). This research examined the change from ICD-9 manual coding to ICD-9 automated coding undertaken at the Australian Bureau of Statistics to assess whether mortality rates were affected by the introduction of ACS. It was found that the mortality rates for several diseases were significantly affected by the introduction of ACS, with pneumonia and senile dementia being the two diseases most affected. With the introduction of ACS, mortality rates for pneumonia were more than doubled and mortality rates for senile dementia were more than halved. These findings are consistent with comparability factors calculated by ABS (ABS, 2000).

Such a discrepancy is due to the different interpretations of ICD rules on causes of death used by manual and automated coding systems (Greig, 1999). Prior to ACS, manual coders used an 'Australian' interpretation of the ICD rules, which tended to be more restrictive than the interpretation used in the American ACS. The general principle and selection rules in ICD-9 are defined by the World Health Organization as follows (WHO, 1993):

General Principle: when more than one condition is entered on the certificate, the condition entered alone on the lowest used line of Part I should be selected only if it could have given rise to all the conditions entered above it.

Rule 1: If the General Principle does not apply and there is a reported sequence terminating in the condition first entered on the certificate, select the originating cause of this sequence. If there is more than one sequence terminating in the condition mentioned first, select the originating cause of the first-mentioned sequence.

Rule 2: If there is no reported sequence terminating in the condition first entered on the certificate, select this first-mentioned condition.

Rule 3: If the condition selected by the *General Principle* or by *Rule 1* or *Rule 2* is obviously a direct consequence of another reported condition, whether in Part I or Part II, select this primary condition.

The difference between the manual coding and ACS specifically relates to the application of the 'Rule 3' and the probable/improbable chains of events leading to death (Greig, 1999). Therefore, the Mortality Reference Group, with representation from the WHO Collaborating Centres (2000), has recommended a series of revisions to Rule 3 developed by the World Health Organisation. The recommended changes clarify the intention of Rule 3, making its interpretation more universal in its application, and thus reducing the discrepancies in mortality data due to different interpretations. The Group noted that the major difficulty with the ICD-9 version of Rule 3 is the extent to which it is open for interpretation, with some countries interpreting it stringently and some interpreting it liberally, thus affecting the corresponding mortality rates. Johansson (1996) indicates that, for internationally comparable data, countries must not only use the same classification system, but also apply the rules for coding in the same way. He suggests that for pneumonia, the coder could interpret Rule 3 in one of two contrasting ways. Firstly, coders could accept the causal sequence which indicates pneumonia as a direct consequence of a wide variety of other documented diseases and injuries. Thus, one of the

other documented conditions might be coded as the underlying cause. Alternatively, the ACS interpretation recognises pneumonia as a direct sequel of a more restricted number of other reported conditions. In such cases, pneumonia is more likely to be chosen as the underlying cause itself. Johansson (1996) recommended that the international comparability of cause of death statistics be improved by using the same coding rules and instructions.

Under the manual system in Australia, coders were strict in their interpretation of Rule 3 in relation to pneumonia. Under ACS, the rule was interpreted more liberally to mean that whenever pneumonia appeared on the death certificate in a causal sequence, it would be more likely to be chosen as the underlying cause of death. Thus, there was a large increase in the number of deaths coded as pneumonia under ACS, and a corresponding fall in diseases that commonly appeared on the death certificate with pneumonia, such as senile dementia (ABS, 2000).

Australia is not the only country to have experienced some issues surrounding the introduction of ACS. Johansson (1996) reported that, in Sweden, the number of deaths coded as pneumonia has increased since ACS was introduced. It was found that a quarter of cases should, in their opinion, have been coded as other underlying cause of death, and another quarter of the deaths should have been attributed to another disease even though the latter was not mentioned on the death certificate. In England and Wales, about 86% of diseases were coded the same using ACS compared with manual coding (Office for National Statistics, 1996).

If changes to mortality statistics are due only to differences in rule interpretation between manual and automated coding, it is postulated that differences in mortality rates should be uniform across different demographic groups, such as age groups and by sex. Nevertheless, this was not the case, with some differences evident for the two major diseases examined: senile dementia and pneumonia. For example, for senile dementia, there was a difference in the percentage of cases coded the same manually and with ACS for males and females (males = 9.6% and females = 14.6%). There was also a difference in the percentage of cases coded the same manually, and with ACS, for different age groups (Figure 5). A similar pattern was observed for pneumonia. This suggests that the effects of the change from manual to automated coding had a differential effect within different demographic groups.

This differential effect may be due largely to differences in disease co-morbidity or coexisting conditions for these different demographic groups. As a result, it is postulated that different combinations of diseases will be reported on the death certificates for different demographic groups, thus influencing the range of diseases that a coder or coding software has to select from when coding an underlying cause of death.

Recently, Anderson et al (2001) examined the impact of the change from automated ICD-9 to automated ICD-10 in America using 1996 data that were dual coded. With this change, mortality data in America were affected significantly by the changing interpretations of ICD rules between these versions. With the introduction of ICD-10, pneumonia rates dropped considerably in American data (due to the change to a more restrictive interpretation of ICD rules in ICD-10 as a result of the work undertaken by the Mortality Reference Group), and other diseases (that would have been attributed previously to pneumonia) rose

considerably. A comparability factor of 0.6982 was reported for pneumonia/influenza between ICD-9 and ICD-10. Similar research needs to be undertaken in Australia to determine these comparability factors and compare them to those reported from the ICD-9 manual to automated dual coding study.

This study may have important implications for coding practices and public health research. It is vital that mortality coders and researchers understand the impact of classification changes on mortality data when they seek to explain and account for variations in mortality trends over time. Analysts and public health researchers need to be aware of what mortality changes represent true changes to the pattern of diseases and what mortality changes are merely artefacts of classification changes. This study indicates that there have been some dramatic changes to mortality trends for some diseases with the introduction of automated coding software. While part of these changes can be adjusted through the use of tools such as comparability factors and back-coding of data, further research is needed to examine whether these adjustments are appropriate and whether such practices will affect long-term mortality trends.

Acknowledgments

We thank the Australian Bureau of Statistics for providing the data used in this study, and the Australian Institute of Health and Welfare and the Australian Bureau of Statistics for their tremendous support of NCCH research.

References

1. Anderson RN, Minino AM, Hoyert DL, Rosenberg HM (2001). Comparability of Cause of Death Between ICD-9 and ICD-10: Preliminary Estimates: National Vital Statistics Reports. Atlanta: Centers for Disease Control and Prevention.
2. Australian Bureau of Statistics (1998). Causes of Death: Australia 1997. Brisbane: Australian Bureau of Statistics.
3. Australian Bureau of Statistics (2000). Causes of Death: Australia 1999. Brisbane: Australian Bureau of Statistics.
4. Greig M (1999). An Australian Experience in Implementing Automated Coding. Brisbane: Australian Bureau of Statistics.
5. Israel RA (1990). Automation of Mortality Data Coding and Processing in the United States of America, *World Health Statistics Quarterly* 43, 259-262.
6. Johansson LA (1996). Need for Clarification of the ICD-10 Mortality Coding Rules. In: Meeting of Heads of WHO Collaborating Centres for the Classification of Diseases, Tokyo, 15-21 October 1996. (Unpublished document WHO/HST/ICD/C/96.18; available on request from Division of Health Situation and Trend Assessment, World Health Organisation, Switzerland).
7. Johansson LA, Ragnar W (2000). Comparing Swedish hospital discharge records with death certificates - implications for mortality statistics, *International Journal of Epidemiology* 29, 495-502.
8. The General Register Officer for Scotland. The introduction of automated cause of death coding. <http://www.gro-scotland.gov.uk>

[/growsweb](#) (retrieved from www on 31 July 2001).

9. Rosenberg HM, Johansson LA, Hoyert DL (2000). Annual Report of the WHO Mortality Reference Group. In: Meeting of Heads of WHO Collaborating Centres for the Classification of Diseases, Brazil, 15-21 October, 2000. (Unpublished document WHO/HST/ICD/C/00.44; available on request from Division of Health Situation and Trend Assessment, World Health Organisation, Switzerland).
10. Taylor R, Lewis M, Powles J (1998). The Australian mortality decline: all-cause mortality 1788-1990, *Australian and New Zealand Journal of Public Health* 22, 27-36.
11. World Health Organisation (1993). International Statistical Classification of Diseases and Related Health Problems (Tenth Revision). Geneva: World Health Organisation.

Authors

All authors are from the National Centre for Classification in Health (NCCH), School of Public Health, Queensland University of Technology, Brisbane, Queensland.

Kirsten McKenzie BSocSc(Hons)(Psych)

Senior Research Assistant, NCCH, Brisbane

Kirsten McKenzie has worked on a number of projects including the examination of the impact of changes in disease classifications and coding practices on long-term mortality trends and mortality data.

Sue Walker BAppSc(MRA), GradDip(Public Health)

Associate Director, NCCH, Brisbane

Sue Walker is a member of the World Health Organization's Mortality Reference Group, as the designate of the Australian Collaborating Centre for the Classification of Diseases.

Shilu Tong BMed, MMed, PhD

Senior Research Fellow, NCCH, Brisbane

Telephone: (+61) 07 3864 9753 Email: s.tong@qut.edu.au

Dr Tong is a medical graduate with a PhD in epidemiology. He has designed and implemented a number of projects including the international comparison of mortality data and the examination of the potential impact of changes in disease classifications and coding practices on long-term mortality trends.



© 2001 Health Information Management Association of Australia Limited