Published by Oxford University Press on behalf of the International Epidemiological Association
© The Author 2007; all rights reserved. Advance Access publication 28 February 2007

International Journal of Epidemiology 2007;**36**:642–651
doi:10.1093/ije/dym003

Validation of cause-of-death statistics in urban China

Chalapati Rao,[1] Gonghuan Yang,[2] Jianping Hu,[3] Jiemin Ma,[2] Wan Xia[2] and Alan D Lopez[1]*

Accepted 22 December 2006

Background National vital registration systems are the principal source of cause specific mortality statistics, and require periodic validation to guide use of their outputs for health policy and programme purposes, and epidemiological research. We report results from a validation of cause of death statistics from health facilities in urban China.

Methods 2917 deaths from health facilities located in six cities in China constituted the study sample. A reference diagnosis of the underlying cause was derived for each death, based on expert review of available medical records, and compared with that filed at registration. Sensitivity, specificity and positive predictive value were computed for specific causes/cause categories according to the International Classification of Diseases (ICD), including analyses based on quality of evidence scores for each cause. Patterns of misclassification by the registration system were studied for individual causes of death.

Results The registration system had good sensitivity in diagnosing cerebrovascular disease and several site specific cancers (lung, liver, stomach, colorectal, breast and pancreas). Sensitivity was average (50–75%) for some major causes of adult death in China, namely ischaemic heart disease (IHD), chronic obstructive lung disease (COPD), diabetes, and liver and kidney diseases, with compensatory misclassification patterns observed between several of them. Sensitivity was particularly low for hypertensive disease.

Conclusions Although diagnostic misclassification is not uncommon in urban death registration data, they appear to balance each other at the population level. Compensating misclassification errors suggest that caution is required when drawing conclusions about particular chronic causes of adult death in China. Investment is required to improve the quality of cause attribution for health facility deaths, and to assess the validity of cause attribution for home deaths. Periodic assessments of the quality of cause of death statistics will enhance their usability for health policy and epidemiological research.

Keywords China, cause of death, mortality statistics, validation, vital registration

Introduction

Cause-of-death statistics are among the most widely used epidemiological data, and are a principal source of information for health policy and planning. Since deaths in most countries must be registered by law, civil registries are an obvious source of mortality data. Following the landmark study of the London 'Bills of Mortality' by John Graunt in 1662 (cited by Greenburg[1]) to guide the control of plague and other epidemics, many countries have developed vital registration systems that yield useful information on causes of death.[2] Nonetheless, registration data potentially suffer from several shortcomings, and require periodic evaluation to assess their reliability.

China is developing a national vital registration system, which currently covers 10% of its population.[3] A recent evaluation of cause of death statistics from the Chinese vital registration system identified the need for validation of reported causes of death.[4] Such validation assessments have important

[1] School of Population Health, University of Queensland Public Health Building, Herston Road, Herston, Qld 4006, Australia.

[2] Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences, 5, Dong Dan San Tiao, 100005 Beijing, China.

[3] Center for Health Statistics Information, Ministry of Health, 1, Nanlu, Xizhimenwai, Xicheng District, Beijing 100044, China.

* Corresponding author. School of Population Health, University of Queensland, Public Health Building, Herston Road, Herston, QLD 4006, Australia. E-mail: a.lopez@sph.uq.edu.au

Downloaded from https://academic.oup.com/ije/article/36/3/642/652082 by guest on 19 April 2024

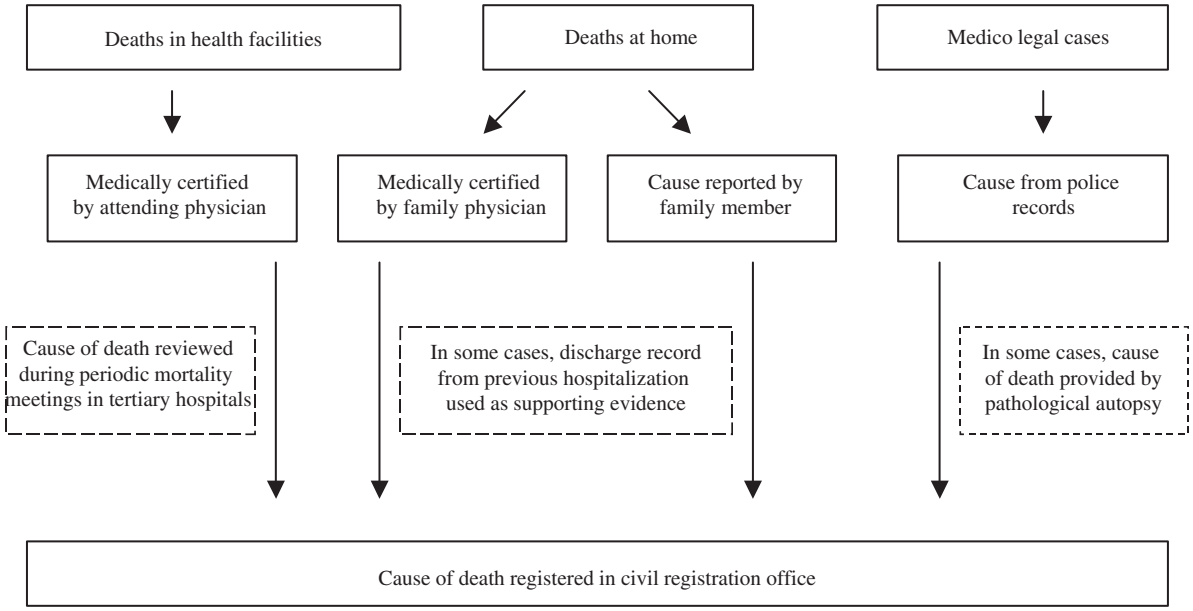

Figure 1 Registration and certification of cause of death in urban China

implications for estimating overall cause-specific mortality in populations,[5,6] and provide an empirical basis to correct registration data.

Validation studies require a 'gold standard' against which to compare the cause of death recorded by the vital registration system. The ideal gold standard causes of death are from autopsy studies,[7–9] but this is often prohibitively expensive. In China, autopsies are usually restricted to medico-legal cases.[10] As an alternative, the next best reference diagnosis for validation is that derived from expert review of high quality medical records.[6,11–13] Given the lack of reliable hospital records in rural China, such validation studies can only be carried out in urban areas. This is, however, of major interest for assessing the reliability of global and regional estimates of causes of death and burden of disease, given the size of the population (>300 million) living in urban areas of China.[14] In this article, we report findings from a study to validate registration diagnoses for major causes of death in six large cities in China.

About 55% of urban deaths occur in health facilities (GH Yang, personal communication), with attending physicians certifying the cause of death (Figure 1) . For home deaths, (37% of urban deaths), some are medically certified by family physicians, using recent hospital discharge records as supporting evidence, although there is uncertainty about the accuracy of discharge diagnoses in representing the illness preceding death. Causes of the remaining domiciliary deaths are reported by family members without physician advice. Diagnostic uncertainty for home deaths is undoubtedly higher than for deaths in health facilities. Under these circumstances, we validated registered causes for deaths that occurred in health facilities only, using underlying causes derived from an expert review of hospital records as reference diagnoses for validation.

There are varying levels of strength of evidence in hospital records. In this study, we have graded the strength of evidence from hospital records for each death using a standard classification system, and assessed validity of registration diagnoses by these evidence categories.

Methods

Study population

The validation study was conducted in six cities; Beijing, Shanghai, Haerbin, Chengdu, Wuhan and Guangzhou, with populations ranging from 7 million (Guangzhou) to 16.5 million (Shanghai). These large cities were chosen as a convenience sample representing urban populations in China, based on the much greater likelihood that medical records for deaths in health facilities in large cities would contain adequate evidence to yield reference diagnoses.

Sampling plan

Resources were available for a total study sample of ~3000 deaths. Cases were distributed across causes of death to approximate to the proportionate distribution of cause-specific mortality as reported in the vital registration system in urban areas for the year 2000,[15] with the following modification. We selectively under-sampled deaths from high frequency chronic causes of adult death such as stroke, ischaemic heart disease (IHD) and chronic obstructive pulmonary disease (COPD), and increased deaths from other causes of public health concern in China, such as selected site-specific cancers, liver diseases, diabetes, infectious diseases, kidney diseases and perinatal conditions (Table 1).

The cases were recruited from city death registers in chronological order during the reference period for the study (June 1 to November 30, 2002). Each city had an approximate target number of cases to be recruited by cause, based on the sampling plan. Additional selection criteria were permanent residency of the deceased, and the occurrence of death in

Downloaded from https://academic.oup.com/ije/article/36/3/642/652082 by guest on 19 April 2024

Table 1 Distribution of study sample by cause group and city, and comparison with proportionate distribution by cause from urban registration data, China 2000

Disease group	ICD Chapter/Codes	City						Total	Registration data (%)	Study sample (%)
		Beijing	Chengdu	Guangzhou	Haerbin	Shanghai	Wuhan			
Infectious diseases	I	2	6	37	3	56	69	173	1.3	5.9
Neoplasms	C00-D48	112	180	146	116	165	119	838	24.4	28.7
Diabetes	E10-E14	29	15	17	15	29	15	120	2.8	4.1
Nervous system	VI	3	3	0	5	21	21	53	2.0	1.8
Cardiovascular diseases	IX	111	98	142	166	138	144	799	40.7	27.4
Respiratory diseases	X	83	89	78	34	41	40	365	13.3	12.5
Digestive diseases	XI	33	24	41	20	18	43	179	3.1	6.1
Kidney diseases	XIV	7	27	19	14	30	21	118	1.5	4.0
Perinatal conditions	XVI	2	21	6	7	10	22	68	0.5	2.3
External causes	XX	12	11	23	10	28	29	113	5.9	3.9
Other diseases	All other codes[a]	12	13	5	8	26	27	91	4.5	3.2
Total		406	487	514	398	562	550	2917	100	100

[a] Chapters V, VII, VIII, XII, XIII, XV, XVII, XVIII and rest of III and IV.

a secondary/tertiary facility equipped with the necessary diagnostic aids.

Data collection and processing

Data was collected from January to December, 2003. For each selected death, details from the death register were computerized, including the ICD code for the underlying cause. Medical records for the admission preceding death were then traced in identified health facilities. Relevant details including the clinical history, examination findings, investigation reports, prescribed treatment, specialist consultant or referral notes and the chronology of events until death were abstracted onto a 'hospital information sheet', along with photocopies of certain laboratory or radiology reports. During case selection, 363 deaths were excluded (11% of the 3280 deaths considered for inclusion) due to insufficient medical records, each being subsequently replaced by an additional death from the register due to the same cause, which met other selection criteria.

A panel of physicians (including an internist, a cardiologist, an oncologist and a respiratory physician) trained specifically in death certification according to ICD-10[16] independently reviewed information sheets and certified causes of death. During training, specific exercises were conducted to assess inter-observer reliability in certification, and ensure common understanding of ICD-10 procedures. Cases were assigned to panel members in accordance with their specialty, with a few additional cases from injuries. Where certification was uncertain for a particular case, it was discussed and resolved by consensus among panel members. Experts at the WHO-ICD Collaborating Centre at Beijing selected and coded an underlying cause from each death certificate, resulting in reference diagnoses for 2917 deaths from the six cities.

Strength of evidence categories were assigned for causes of death listed on each certificate, according to the criteria shown in Table 2. These criteria are similar to those proposed by Moriyama and colleagues who graded strength of evidence as 'very good', 'good' or 'sketchy'.[17]

Table 2 Categories for recording strength of evidence from medical records

Code	Evidence category	Criteria
1	Weak	Clinical diagnosis based on history/physical examination
2	Probable	Imaging, biochemistry and electrocardiograph
3	Confirmed	Autopsy, histopathology, microbiology and operative findings

For instance, in the case of ischaemic heart disease, the strength of evidence would progressively increase if the diagnosis is made on the basis of clinical symptoms alone, or was additionally supported by an electrocardiogram, evidence from imaging procedures, or from operative findings, or from an autopsy. The criteria could also vary for different diagnoses; e.g. clinical features are more helpful than radiographs in diagnosing pneumonia at early ages,[18] imaging provides adequate *ante mortem* evidence for cerebrovascular disease and biochemical or serological tests have a small margin of error, which could affect the strength of evidence for certain causes. While admittedly somewhat arbitrary, these criteria preserve a commonsense hierarchy of confidence associated with varying forms of medical evidence.

Statistical analyses

Underlying causes from each source (registration and reference diagnoses) were aggregated as per the ICD Mortality Tabulation List 1,[19] for primary validation analysis by cause, in terms of sensitivity, specificity and positive predictive value of the registration system. Similar analyses were conducted for different levels of strength of evidence for each cause category in the ICD Mortality List 1. In addition, we measured validity for certain three-character ICD codes to evaluate diagnostic accuracy for specific causes likely to be of interest for epidemiological research.[20]

Downloaded from https://academic.oup.com/ije/article/36/3/642/652082 by guest on 19 April 2024

Results

Measures of validity

Table 3 gives an overview of the validation results, for all grades of evidence combined, with causes listed in descending order as reported by the routine system. Overall, the findings on sensitivity suggest that the system appears to be functioning reasonably satisfactorily, with sensitivity for 11 leading cause categories exceeding 75% (good sensitivity). However, the routine system was less successful (average sensitivity; 50–75%) in identifying deaths due to some important non-communicable diseases in China, namely IHD, COPD, diabetes, liver and kidney diseases, and was particularly poor (sensitivity <50%) in correctly identifying deaths from hypertensive disease.

Although not a measure of validity as such, positive predictive value is useful as an operational measure of the reliability of the registration system, as it reflects both test validity in terms of sensitivity as well as the proportion of deaths in the sample due to the cause of interest As in the analysis of sensitivity, hypertensive diseases and pneumonia also scored poorly in terms of PPV (<50%), and the details of misclassification involving these diseases are described in Table 4. However, the low proportions of these two conditions could have lowered the PPV. Interestingly, the PPV for COPD was reasonably reliable (78%), probably influenced by its high proportionate representation. This is important since China is estimated to account for about half of global mortality from COPD.[21]

The net changes to crude cause-specific mortality fractions for urban China implied by these findings are not substantial, being relatively minor (<10%) for some leading causes of death such as stroke, COPD and several site specific cancers. However, the implied changes for some other diseases of major public health importance are more

Table 3 Validation characteristics for 25 causes of death in urban China, 2002

		Numbers of deaths							
				Reassigned					
Cause of death	**ICD Codes**	In routine registration	Confirmed by panel	To other causes	From other causes	Final total	**Sensitivity (95% CI)**	**PPV (95% CI)**	**Change in cause-specific mortality fraction (%)[a]**
Cerebrovascular diseases	I60-I69	477	422	55	95	517	81.6 (78–85)	88.4 (86–91)	−8.4
Lung cancer	C34	280	266	14	15	281	94.7 (92–97)	95.0 (93–98)	−0.4
Ischaemic heart disease	I20-I25	231	195	36	86	281	69.4 (64–75)	84.4 (79–89)	−21.6
Chronic obstructive pulmonary disease	J40-J47	228	178	50	64	242	73.5 (68–79)	78.1 (72–83)	−6.1
Liver cancer	C22	148	133	15	18	151	88.1 (83–93)	89.9 (85–95)	−2.0
Diabetes mellitus	E10-E14	120	65	55	39	104	62.5 (53–72)	54.1 (45–63)	13.3
Diseases of the genitourinary system	N00-N98	118	36	82	24	60	60.0 (46–71)	30.5 (21–44)	49.2
Stomach cancer	C16	102	99	3	9	108	91.7 (86–97)	97.1 (94–100)	−5.9
Diseases of the liver	K70-K76	98	56	42	20	76	73.6 (64–84)	57.1 (47–66)	22.4
Viral hepatitis	B15-B19	89	72	17	42	114	63.1 (54–72)	80.9 (72–89)	−28.1
Cancer of colon, rectum and anus	C18-C21	85	83	2	14	97	85.5 (79–93)	97.7 (94–100)	−14.1
Pneumonia	J12-J18	76	11	65	9	20	56.0 (33–76)	14.4 (7–22)	73.7
Perinatal conditions	P00-P96	68	68	0	6	74	91.9 (85–98)	100.0	−8.8
Respiratory tuberculosis	A15-A16	67	58	9	9	67	86.5 (78–95)	86.5 (78–95)	0.0
Falls	W00-W19	39	25	14	9	34	73.5 (59–88)	64.1 (49–79)	12.8
Oesophageal cancer	C15	33	31	2	0	31	100.0	93.9 (86–100)	6.1
Breast cancer	C50	32	31	1	3	34	91.2 (82–100)	96.9 (91–100)	−6.3
Transport accidents	V01-V99	31	30	1	6	36	83.3 (71–95)	96.7 (90–100)	−16.1
Anaemias	D50-D64	32	21	11	3	24	87.5 (74–100)	65.6 (49–82)	25.0
Congenital malformations	Q00-Q99	31	26	5	9	35	74.3 (60–89)	83.9 (71–97)	−12.9
Rheumatic heart disease	I00-I09	29	24	5	9	33	72.7 (57–87)	82.7 (69–96)	−13.8
Pancreatic cancer	C25	28	25	3	2	27	92.6 (83–100)	89.3 (78–100)	3.6
All other external causes	W20-Y89	26	11	15	25	36	30.6 (16–46)	42.3 (23–63)	−38.5
Hypertensive disease	I10-I15	26	11	15	50	61	18.1 (8–27)	42.3 (23–61)	−134.6
Gastric and duodenal ulcer	K25-K27	16	11	5	10	21	52.4 (31–74)	68.8 (46–91)	−31.3
All other causes	All other codes	407				353			
Total deaths		2917				2917			

[a] Positive change indicates overestimate by routine system, negative change indicates underestimate.

Downloaded from https://academic.oup.com/ije/article/36/3/642/652082 by guest on 19 April 2024

Downloaded from https://academic.oup.com/ije/article/36/3/642/652082 by guest on 19 April 2024

Table 4 Misclassification matrix for communicable and non-communicable diseases, urban China, 2002

	Medical records diagnoses														
Registration diagnoses	Cerebro-vascular diseases	IHD	Rheumatic heart disease	Hyper-tensive diseases	COPD	Pneumonia	Other respiratory diseases[b]	Diabetes Mellitus	Genito-urinary diseases	Viral hepatitis	Gastric and duodenal ulcer	Diseases of the liver	Other digestive diseases[c]	All other diseases	**Total registration deaths**
Cerebrovascular diseases	**422**	12	4	6	4			10	4		1		3	11	477
Ischaemic heart disease	13	**195**		4	6		2	2	1				2	6	231
Rheumatic heart disease	2	3	24												29
Hypertensive diseases	5	3	1	11			1	3	1					1	26
Other heart diseases[a]	4	8	2	3	9		2						1	7	36
COPD	7	9	1	3	**178**	3	5	2	4				4	12	228
Pneumonia	10	15		3	15	**11**	7	6					2	7	76
Other respiratory diseases[b]	8	6		1	5	6	**18**	4			1	1	2	8	60
Nervous system diseases	8	3		1	2		4	1						34	53
Diabetes Mellitus	17	13		3	5		1	**65**	1				6	9	120
Genitourinary diseases	6	5		23	5		2	10	**45**		1	2	2	17	118
Viral hepatitis										**72**		8		9	89
Gastric and duodenal ulcer	1										11	2	1	1	16
Diseases of the liver										38	1	**56**		3	98
Other digestive diseases[c]	1	1		1			2	1	1		2	5	**42**	7	63
All other diseases	13	8	1	2	13		3	0	3	4	4	2	7	1137	1197
Total Medical Records deaths	517	281	33	61	242	20	47	104	60	114	21	76	72	1269	2917

Values in bold indicate matched cases in the cross tabulation for each diagnoses from the two sources.

[a] 126–151.

[b] J00–J06, J30–J39, J60–J98.

[c] K00-K22, K28–K66, K80–K92.

significant, including IHD and viral hepatitis (undercount in vital registration), and diabetes, pneumonia, and genitourinary and liver diseases (overcount).

For some purposes, particularly epidemiological research into the causes of various specific diseases, greater detail about cause specific reliability may be required. For instance, while the sensitivity for cerebrovascular diseases (ICD codes I60–I69) was 82%, that for its three major sub categories, which are of interest for specific epidemiological enquiry, was much lower, as follows:

- intracerebral haemorrhage (I61): 69%
- cerebral infarction (I63): 66%
- sub arachnoid haemorrhage (I60): 60%

Sensitivity was also low for components of IHD, diabetes and COPD.

Misclassification

A key outcome of validation studies is to identify, for any given condition, the principal diseases and/or injuries involved in its misclassification. The principal misclassification patterns for major causes appear largely compensatory (Table 4), although important age and sex distortions may still occur. The table also suggests that the principal reasons for the undercount of ischaemic heart disease in official Chinese mortality data is misclassification to stroke, diabetes, pneumonia and other forms of heart disease. Other important misclassification patterns include the systematic coding of hepatitis deaths to other liver diseases, excessive coding of pneumonia among respiratory disease deaths, and the poor distinction between hypertensive diseases and genitourinary system disorders. Of particular concern is COPD, given the role that mortality in China is thought to play in global estimates. The 50 misclassified COPD deaths in the routine system are distributed across a wide range of causes, and these findings suggest that estimates of disease frequency need to be viewed carefully and in close association with vascular disease mortality.

Although the number of deaths in the sample from external causes is small, there appears to be some confusion in cause of death attribution. Over 12% (14) of reported injury deaths (mostly falls) were attributed to other (natural) causes on the basis of medical records. However, the overall proportionate distribution of mortality across diseases and injuries may not be that inaccurate since 20 deaths reported as due to natural causes by the routine system were reclassified as injuries by the expert review.

Quality of medical evidence

The degree of confidence in such an evaluation depends on the reliability of the medical expert diagnoses, which in turn is strongly dependent on the nature and precision of the available clinical evidence. Table 5 summarizes the distribution of cases according to strength of evidence category. Among neoplasms, lung, oesophagus, stomach and colorectal cancers had several cases with confirmed diagnoses, but this was not so for breast and liver cancer deaths, despite the apparent ease in obtaining a tissue sample from these organs. While the highest grade of strength of evidence is not common for several conditions (e.g. IHD, stroke, tuberculosis and nephritis), evidence for a probable diagnosis is sufficient to yield reasonable confidence in the quality of the reference diagnosis. This is less obvious, however, for COPD and pneumonia, with several cases of each in the reference data set being based on 'weak' evidence. The relatively poor validation characteristics for these diseases need to be viewed with this in mind.

On recomputing the validation parameters using only cases with 'confirmatory' or 'probable' evidence, we observed that sensitivity is higher when measured against more rigorous evidence for several leading causes of death, as might be expected (Figure 2). This is not the case for COPD, however, for which a diagnosis based on weak evidence may not be very meaningful, and for perinatal conditions, the diagnosis of which, given the strong age dependence, is largely unaffected by strength of evidence. The wide confidence intervals for IHD and diabetes, when assessed against weak evidence, reflect the small number of cases in this category.

For cancers, a more appropriate comparison would be to assess sensitivity of diagnosis based on pathology against the other categories combined (Figure 3). Again, sensitivity is greater when assessed against the highest standards of clinical evidence, but only marginally so.

Table 5 Strength of evidence frequencies for selected causes of death in urban China, 2002

Cause of death	Deaths (number) assessed from medical records	Confirmed (%)	Probable (%)	Weak (%)
Cerebrovascular disease	517	0	86	14
Lung cancer	281	41	49	10
Ischaemic heart disease	281	0	98	2
Chronic obstructive lung disease	242	0	69	31
Liver cancer	151	1	89	10
Stomach cancer	108	30	54	16
Viral hepatitis	114	0	100	0
Diabetes mellitus	104	0	94	6
Colorectal cancer	97	23	58	18
Liver diseases	76	19	67	15
Perinatal conditions	74	1	45	53
Tuberculosis	67	0	92	8
Hypertensive diseases	61	0	98	2
Pneumonia	20	0	77	23
Other malignant neoplasms	50	50	2	48
Congenital malformations	35	9	31	60
Breast cancer	34	6	94	0
Genitourinary diseases	60	6	78	22
Rheumatic heart disease	33	0	97	3
Oesophageal cancer	31	32	52	16

Downloaded from https://academic.oup.com/ije/article/36/3/642/652082 by guest on 19 April 2024

Discussion

To the extent that cause of death information is being used to inform policy debates and to guide priority setting in the health sector, it is important to know their quality at the population level. This is particularly true in populations where not all deaths occur in health establishments. In this study, we have evaluated, for the first time, the validity of cause attribution for deaths occuring in health facilities in urban China. Findings from a study that assessed the reliability (but not validity) of rural cause of death data are reported elsewhere.[22]

The basic design of the study, namely to compare diagnoses from the routine vital registration system with a reference diagnosis based on expert review of best available clinical evidence, appears to work satisfactorily in many different settings. For example, Gittelsohn *et al*.[23] compared causes of death on death certificates with diagnoses recorded on hospital records for 9724 deaths in Vermont State, USA (1969–75), and found the agreement at ICD-9 3-character level to be no more

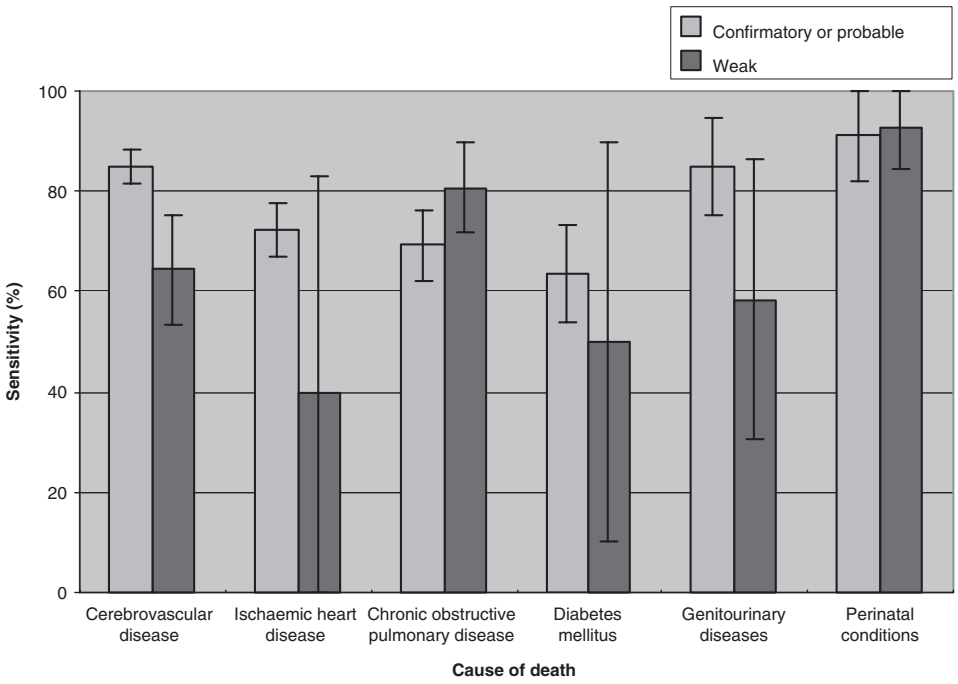

Figure 2 Variations in sensitivity by strength of evidence for selected causes of death in urban China, 2002

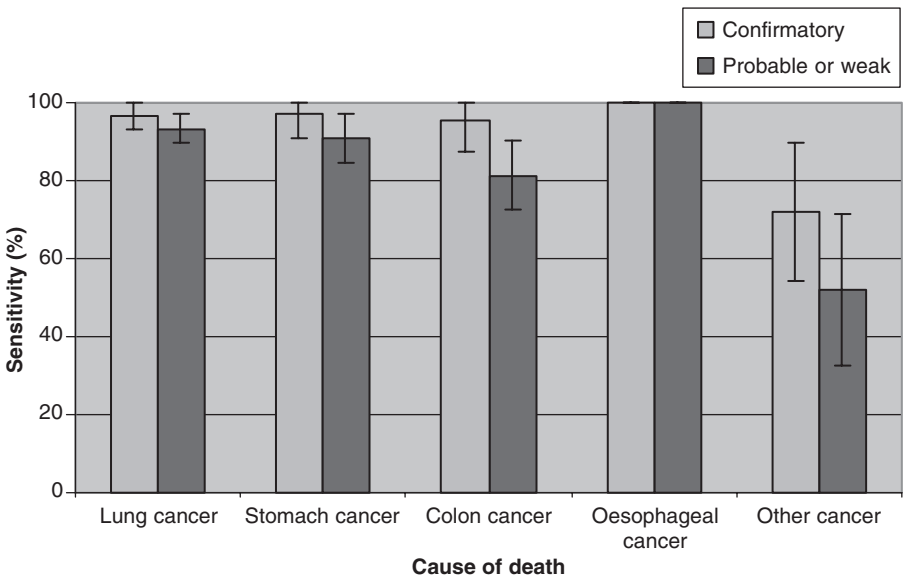

Figure 3 Variations in sensitivity by strength of evidence for cancers, urban China, 2002

Downloaded from https://academic.oup.com/ije/article/36/3/642/652082 by guest on 19 April 2024

than 72%. Johansson and Westerling[12] compared diagnoses on death certificates and hospital discharge records for 69 818 people who had been hospitalized in Sweden during their final year of life in 1995, and found that adding hospital discharge data would have changed the underlying cause of death in 11% of cases. Kuller *et al*.[24] compared death certificates and clinical records for a sample of 16 956 deaths in USA in 1965, and found the sensitivity of stroke diagnoses on death certificates to be 58%. A comparison study between death certificate diagnoses and cancer registry records for 7146 deaths in Saskatchewan, Canada[25] found that stomach, lung and pancreatic cancers were overdiagnosed, and breast cancer underdiagnosed. Interestingly, a comparison between death certificates and tuberculosis register cards for 108 cases in the USA during 1966–67[26] reported that in 46% of cases, the death certificate diagnosis of tuberculosis was open to question.

We believe that the review of medical records is an appropriate procedure, given the significant advancement in *ante-mortem* diagnostic ability using sophisticated imaging and laboratory technology. The accuracy of diagnoses contained in medical records is likely to have improved in recent times, although this development does not replace autopsies as the 'gold' standard in determining the cause of death.[27] However, as autopsy rates fall worldwide,[28] there will be an increasing need to rely on evidence from hospital records to establish reference diagnoses, and greater confidence in the findings of validation studies will be possible provided information on the strength of evidence is routinely available.

Although the number of cases was constrained by costs, the 95% CIs provide some reassurance about the adequacy of our sample size. Indeed, the 95% CI around point estimates of sensitivity for individual causes fell entirely within the threshold margins for each category of sensitivity as follows: good sensitivity (10 causes); average sensitivity (three) and poor sensitivity (two), i.e. for a total of 15 out of 25 reported causes. For four other causes, the width of the 95% CI did not exceed 25%, which is the width of each sensitivity category. The wide 95% CI for sensitivity of pneumonia is not unexpected, given known problems with certification and coding of this condition.[29] Although greater sample size would have enabled more precision in our findings, the results are unlikely to be greatly affected by the number of cases we had recruited in our study.

The 363 deaths that were excluded due to inadequate information were mostly deaths from cardiovascular causes and COPD, with a few each due to diabetes, kidney diseases and some site specific cancers. The replacement strategy used (see 'Methods') in order to achieve the targeted proportionate cause distribution in the sample could potentially introduce some bias in the quality of evidence analysis, since the replacement case could have been assigned to any of the three evidence categories. However, we feel that this was unlikely to be substantial, and in any case is impossible to measure in terms of the extent or direction of this bias.

Another possible source of bias lies in the strategy used to selectively under or over sample deaths from specific diagnostic categories. This decision was based on the premise that a minimum number of deaths from each cause to understand the parameters of diagnostic validity, as well as patterns of misclassification by cause. This could potentially affect the generalizability of the reported changes in cause-specific mortality fractions, described in Table 3.

In comparing diagnoses from the two sources (registration vs medical records review) on a case-by-case basis, disagreement in the underlying cause could be real, or an artefact of differences in certification and coding practices. While training medical experts in death certification from medical records, we conducted assessments of reliability, and found that identification of reference diagnoses was unequivocal in cases with adequate clinical evidence in the hospital records. In implementing the study, in those cases (~15%) with somewhat vague clinical evidence, the reference diagnosis was determined by consensus. These cases were classified in the 'weak' evidence category. We adopted this pragmatic approach to mimic common clinical practice, where in doubtful situations, second or consensus opinions are sought to arrive at a specific diagnosis. This uncertainty could be reduced by more detailed documentation of clinical histories.

Our analyses and results are based on comparisons of underlying causes alone, in order to be able to assess the quality of data from the routine registration system, which reports statistics based on underlying causes only. Further analyses of multiple causes of death could identify discrepancies in certification or coding procedures, and help explain the misclassification patterns for individual causes, as reported elsewhere.[11] Such analyses could also possibly explain the lower sensitivity scores for specific components of broad disease categories (e.g. cerebrovascular disease, COPD and IHD). Under these circumstances, epidemiological or clinical intervention studies in China, which use registration data to determine these specific disease end-points, should be interpreted with great caution.

The study results suggest that cause of death data from health facilities in urban China are prone to misclassification. Compensating misclassification errors do not inspire confidence in the procedures applied to develop individual diagnoses, but appear to balance each other at the population level, a finding observed in dual certification studies elsewhere.[30–32] There are a number of exceptions however, particularly for leading chronic causes of adult death such as IHD, COPD and diabetes. The underdiagnosis of deaths from IHD is typical of the pattern in East Asia,[21] and is a major concern for studies of the epidemiological transition in the region. Even marginal misclassification effects for these causes might have implications for understanding the comparative importance of diseases in China, and the factors that may cause them. Moreover, relying on compensatory patterns of misclassification to yield reasonable descriptions of underlying disease patterns is hazardous and opportunistic, and may not always apply, particularly if diagnostic 'fashions' in cause of death certification and coding, or other diagnostic influences, are operating to systematically bias diagnoses. Any systematic sampling biases in such validation exercises will exacerbate these effects, and restrict the generalizability of the findings. Given that this study has only assessed cause attribution for facility based deaths, the findings reported here probably represent a 'best case' scenario for the quality of urban mortality statistics in

Downloaded from https://academic.oup.com/ije/article/36/3/642/652082 by guest on 19 April 2024

China. Further studies to validate registered causes for the 37% of deaths that occur at home, either using hospital discharge records or verbal autopsy procedures, would be useful in developing more accurate assessments of the validity of cause of death statistics in urban China. There is a clear need to improve cause attribution on a case by case basis, to ensure more reliable population level statistics.

Periodic evaluations of the type reported here have several advantages. They provide direct evidence about the extent and direction of cause of death misclassification and hence enhance the utility of the evidence base for policy action. They also create awareness about the potential utility of cause of death data for health development, particularly in countries developing vital registration systems. Such evaluations also help to identify systemic problems with maintenance of medical records and death certification/coding procedures, and suggest priorities to build capacity to strengthen these critical functions of vital registration.

Acknowledgement

The authors acknowledge financial support for this project from the US National Institute on Aging (Research Grant No. PO1 AG17625). The funding agency had no role in the study design, data collection and analysis or in the preparation of this manuscript.

Conflict of interest: None declared.

KEY MESSAGES

- National cause of death statistics from vital registration systems should be validated periodically using best available reference diagnoses.
- Evaluation of quality of evidence in medical records is important in interpreting findings from validation studies.
- Observed misclassification patterns help understand biases in registration data, and guide their interpretation and use for health policy and programme evaluation.
- Compensating misclassification errors affect the utility of urban Chinese registration data for epidemiological research.

References

[1] Greenberg SJ. The 'Dreadful Visitation': public health and public awareness in seventeenth-century London. *Bull Med Libr Assoc* 1997;**85**:391–401.

[2] Mathers CD, Ma Fat D, Inoue M *et al*. Counting the dead and what they died of: an assessment of the global status of cause of death data. *Bull World Health Organ* 2005;**83**:171–77.

[3] Yang G, Hu J, Rao KQ *et al*. Mortality registration and surveillance in China: History, current situation and challenges. *Popul Health Metr* 2005;**3**:3.

[4] Rao C, Lopez AD, Yang GH *et al*. Evaluating national cause of death statistics: principles and application to the case of China. *Bull World Health Organ* 2005;**83**:618–25.

[5] Johansson LA, Westerling R. Comparing Swedish hospital discharge records with death certificates: implications for mortality statistics. *Int J Epidemiol* 2000;**29**:495–502.

[6] Lahti RA, Penttila A. The validity of death certificates: routine validation of death certification and its effects on mortality statistics. *Forensic Sci Int* 2001;**115**:15–32.

[7] Shojania KG, Burton EC, McDonald KM *et al*. Changes in rates of autopsy-detected diagnostic errors over time: a systematic review. *JAMA* 2003;**289**:2849–56.

[8] Haheim LL. Validation of causes of death by age. *Tidsskr Nor Laegeforen* 1999;**119**:826–30.

[9] Saito I. Review of death certificate diagnosis of coronary heart disease and heart failure in Japan. *Nippon Koshu Eisei Zasshi* 2004;**51**:909–16.

[10] Cao L. [Autopsy practised in modern China]. *Zhonghua Yi Shi Za Zhi* 1994;**24**:154–57.

[11] Alderson MR, Meade TW. Accuracy of diagnosis on death certificates compared with that in hospital records. *Br J Prev Soc Med* 1967;**21**:22–29.

[12] Johansson LA, Westerling R. Comparing hospital discharge records with death certificates: can the differences be explained? *J Epidemiol Community Health* 2002;**56**:301–8.

[13] Nashelsky MB, Lawrence CH. Accuracy of cause of death determination without forensic autopsy examination. *Am J Forensic Med Pathol* 2003;**24**:313–19.

[14] United Nations Population Division. *World Population Prospects: The 2002 Revision*. Vols I–III. New York: United Nations, 2003.

[15] Ministry of Health China. *Annual Internal Report*. Beijing: Ministry of Health, P.R. China, 2001.

[16] World Health Organization. Mortality: guidelines for certification and rules for coding. *International Statistical Classification of Diseases and Health Related Problems—Tenth Revision. Volume 2: Instruction Manual*. Geneva: World Health Organization, 1993; 30–65.

[17] Moriyama IM, Baum WS, Haenszel WM *et al*. Inquiry into diagnostic evidence supporting medical certifications of death. *Am J Public Health* 1958;**48**:1376–87.

[18] British Thoracic Society of Standards of Care Committee. BTS Guidelines for the Management of Community Acquired Pneumonia in Childhood. *Thorax* 2002;**57**:1i–24.

[19] World Health Organization. Special tabulation lists for mortality and morbidity; Mortality tabulation list 1. *International Statistical Classification of Diseases and Health Related Problems—Tenth Revision. Volume 1*. Geneva: World Health Organization, 1993; 1207–10.

[20] Winthereik BR. 'We fill in our working understanding': on codes, classifications and the production of accurate data. *Methods Inf Med* 2003;**42**:489–96.

[21] Murray CJL, Lopez AD. *The global burden of disease: a comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020*. Cambridge, MA: Published by the Harvard School of Public Health on behalf of the World Health Organization and the World Bank; Distributed by Harvard University Press, 1996.

Downloaded from https://academic.oup.com/ije/article/36/3/642/652082 by guest on 19 April 2024

[22] Wang L, Yang G, Ma J *et al*. Evaluation of the quality of cause of death statistics in rural China using verbal autopsies. *J Epidemiol Commun Health* 2006; (In press).

[23] Gittelsohn A, Senning J. Studies on the reliability of vital and health records: I. Comparison of cause of death and hospital record diagnoses. *Am J Public Health* 1979;**69**:680–89.

[24] Kuller LH, Bolker A, Saslaw MS *et al*. Nationwide cerebrovascular disease mortality study. II. Comparison of clinical records and death certificates. *Am J Epidemiol* 1969;**90**:545–55.

[25] Barclay TH, Phillips AJ. The accuracy of cancer diagnosis on death certificates. *Cancer* 1962;**15**:5–9.

[26] Khoury SA. Death certificates and tuberculosis register cards. A correlation study of 108 cases. *Am Rev Respir Dis* 1971;**104**:936–37.

[27] Autopsy. A comprehensive review of current issues. Council on Scientific Affairs. *JAMA* 1987;**258**:364–69.

[28] Hill RB, Anderson RE. The recent history of the autopsy. *Arch Pathol Lab Med* 1996;**120**:702–12.

[29] Goldacre MJ, Duncan M, Cook-Mozaffari P *et al*. Mortality rates for common respiratory diseases in an English population 1979-1998: artefact and substantive trends. *J Public Health* 2004;**26**:8–12.

[30] Heaseman M, Lipworth L. Accuracy of certification of cause of death: Studies on Medical and Population Subjects, No 20. London: Her Majesty's Stationary Office, 1966.

[31] Moriyama IM, Dawber TR, Kannel WB. Evaluation of diagnostic information supporting medical certification of deaths from cardiovascular disease. In: Haenzel W (ed.). *Epidemiological Approaches to the Study of Cancer and Other Chronic Diseases*. Washington: Government Printing Office, 1966.

[32] James G, Patton RE, Heslin AS. Accuracy of cause-of-death statements on death certificates. *Public Health Rep* 1955;**70**:39–51.

The online version of this article has been published under an open access model. Users are entitled to use, reproduce, disseminate, or display the open access version of this article for non-commercial purposes provided that: the original authorship is properly and fully attributed; the Journal and Oxford University Press are attributed as the original place of publication with the correct citation details given; if an article is subsequently reproduced or disseminated not in its entirety but only in part or as a derivative work this must be clearly indicated. For commercial re-use, please contact journals.permissions@oxfordjournals.org
Published by Oxford University Press on behalf of the International Epidemiological Association *International Journal of Epidemiology* 2007;**36**:651–653
© The Author 2007; all rights reserved. Advance Access publication 24 May 2007
doi:10.1093/ije/dym108

Commentary: Reliable measurement of the causes of mortality in developing countries

Prabhat Jha,[1]* Binu Jacob[1] and Rajesh Kumar[2]

Public health in industrialized countries was transformed when vital statistics on age, sex and socioeconomic distribution of deaths by cause became available in the 19th and 20th centuries.[1] These statistics have shown good news, such as the large declines in under-5 mortality and tuberculosis mortality during the 20th century. They have also raised alarm; in the mid 1940s, a dramatic increase in lung cancer deaths in British and American men after World War II led to a great deal of research on smoking.[2] In the early 1980s, routine mortality data from San Francisco revealed an exceptional increase in immune-related deaths among young men and signalled the start of the American HIV-1 epidemic.[3] Routinely collected data have helped to spur further research and public health action and contributed to the enormous increases in life expectancy in the 20th century.[4]

About 46 million of the estimated 60 million deaths worldwide occur in developing countries, where death registration and medical attention at the time of death is low. A recent review of 115 countries that report mortality to the World Health Organization (WHO) found that only 64 had complete death registration with good quality and coverage of cause of deaths. Seventy-five countries, including 90% of those in the African region, did not provide data on causes of death for any year after 1990.[5]

In this issue, Chalapati Rao and colleagues[6] present important new research that examines the validity of causes of death in urban China from routine registration. They compare about 3000 deaths that occurred in urban health facilities with detailed hospital records. They find that the routine registration system has reasonably high sensitivity versus hospital records for some of the major causes of death in China (such as stroke, specific cancers), and only modest sensitivity for some other leading killers (such as chronic lung disease and heart attack). Importantly, they find that specificity is reasonably high.

What then are the implications of their findings for China, and for other developing countries? We think there are three. Note that the implications differ for capturing the *act* of death and for documenting *causes* of death.

First, routine death data is likely to be useful to monitor future trends in mortality by cause in urban China. The coverage of routine death registration has been stable for most of the last decade, but should expand in the future. Notwithstanding the misclassification of causes from routine

[1] Centre for Global Health Research, Li Ka Shing Knowledge Institute, St. Michael's Hospital and Public Health Sciences, University of Toronto, Toronto, Canada.

[2] School of Public Health, Post Graduate Institute of Medical Education and Research, Chandigarh, India.

* Corresponding author. E-mail: prabhat.jha@utoronto.ca

Downloaded from https://academic.oup.com/ije/article/36/3/642/652082 by guest on 19 April 2024