Validation of the Symptom Pattern Method for Analyzing Verbal Autopsy Data

Christopher J. L. Murray[1*], Alan D. Lopez[2], Dennis M. Feehan[1], Shanon T. Peter[3], Gonghuan Yang[4]

1 Institute for Health Metrics and Evaluation, University of Washington, Seattle, Washington, United States of America, 2 School of Population Health, University of Queensland, Herston, Queensland, Australia, 3 Harvard Initiative for Global Health, Cambridge, Massachusetts, United States of America, 4 China Center for Disease Control and Prevention, Beijing, China

Abbreviations: CSMF, cause-specific mortality fraction; ICD, International Classification of Diseases; PCVA, physician-coded verbal autopsy; SEAR-D, WHO South East Asia subregion D; SP, symptom property method; VA, verbal autopsy; WHO, World Health Organization; WPR-B, WHO Western Pacific subregion B

* To whom correspondence should be addressed. E-mail: cjlm@u.washington.edu

ABSTRACT

Background

Cause of death data are a critical input to formulating good public health policy. In the absence of reliable vital registration data, information collected after death from household members, called verbal autopsy (VA), is commonly used to study causes of death. VA data are usually analyzed by physician-coded verbal autopsy (PCVA). PCVA is expensive and its comparability across regions is questionable. Nearly all validation studies of PCVA have allowed physicians access to information collected from the household members' recall of medical records or contact with health services, thus exaggerating accuracy of PCVA in communities where few deaths had any interaction with the health system. In this study we develop and validate a statistical strategy for analyzing VA data that overcomes the limitations of PCVA.

Methods and Findings

We propose and validate a method that combines the advantages of methods proposed by King and Lu, and Byass, which we term the symptom pattern (SP) method. The SP method uses two sources of VA data. First, it requires a dataset for which we know the true cause of death, but which need not be representative of the population of interest; this dataset might come from deaths that occur in a hospital. The SP method can then be applied to a second VA sample that is representative of the population of interest. From the hospital data we compute the properties of each symptom; that is, the probability of responding yes to each symptom, given the true cause of death. These symptom properties allow us first to estimate the population-level cause-specific mortality fractions (CSMFs), and to then use the CSMFs as an input in assigning a cause of death to each individual VA response. Finally, we use our individual cause-of-death assignments to refine our population-level CSMF estimates. The results from applying our method to data collected in China are promising. At the population level, SP estimates the CSMFs with 16% average relative error and 0.7% average absolute error, while PCVA results in 27% average relative error and 1.1% average absolute error. At the individual level, SP assigns the correct cause of death in 83% of the cases, while PCVA does so for 69% of the cases. We also compare the results of SP and PCVA when both methods have restricted access to the information from the medical record recall section of the VA instrument. At the population level, without medical record recall, the SP method estimates the CSMFs with 14% average relative error and 0.6% average absolute error, while PCVA results in 70% average relative error and 3.2% average absolute error. For individual estimates without medical record recall, SP assigns the correct cause of death in 78% of cases, while PCVA does so for 38% of cases.

Conclusions

Our results from the data collected in China suggest that the SP method outperforms PCVA, both at the population and especially at the individual level. Further study is needed on additional VA datasets in order to continue validation of the method, and to understand how the symptom properties vary as a function of culture, language, and other factors. Our results also suggest that PCVA relies heavily on household recall of medical records and related information, limiting its applicability in low-resource settings. SP does not require that additional information to adequately estimate causes of death.

The Editors' Summary of this article follows the references.

Introduction

Cause of death data are a critical input into debates on public health, assessment of epidemiological patterns and trends, and guidance on allocation of scarce resources for pubic health and medical care [1]. Cause of death data for these purposes are equally important for high-income and low-income countries, but the challenge of comparable measurement is different. For high-income countries, the main task is to minimize systematic differences in assignment of causes of death over time and across populations. For developing countries, where vital registration systems are generally not complete, and progress toward higher levels of coverage and certification has been unacceptably slow, the challenge is how to make better use of alternative cause-of-death measurement approaches to reliably estimate population-level mortality patterns. This paper presents a new method for assigning causes of death based on an interview of members of the decedent's household.

Considerable effort has been invested to assess mortality by cause based on information collected after death from household members—termed since the early 1980s "verbal autopsy." Use of this approach can be traced to 1931 [2], but the first large-scale systematic application of it was in Matlab, Bangladesh [3]. Because vital registration coverage has not significantly improved in developing countries, interest in verbal autopsy (VA) methods has expanded to at least four distinct settings: clinical trials and large-scale epidemiological studies; demographic surveillance systems, such as those in the INDEPTH network [4]; national sample surveillance systems [5,6]; and household surveys [2]. Despite increasing and widespread use of VA in field data collection, utilization of VA data for national epidemiological monitoring and global/regional burden of disease estimation has so far been limited. For example, WHO makes almost no use of VA data for adult cause of death estimation because of the heterogeneity of VA implementation and non-ICD (International Classification of Diseases) cause lists. A notable exception has been the use of VA data (for under-age-5 cause of death) to dramatically reduce global estimates of measles mortality in children [7].

The heterogeneity of VA design and implementation spans several critical issues: instrument design, method for assigning underlying cause of death, cause of death lists, recall period, and type of respondent [8]. Two are particularly critical: instrument design and method for cause-of-death assignment. First, the instrument used to collect data after death from a household member varies widely, limiting comparability. Some field sites use open-ended questions in an unstructured interview, while other instruments are highly structured checklists of symptoms and signs. Most instruments also elicit information on health-care use, medical diagnoses, and documentation of the cause of death. Second, data collected through the VA instrument are analyzed to assign an underlying cause of death in one of three ways: physician review, expert algorithms, and statistical models. Physician review is the dominant approach, since the critical review of signs and symptoms approximates the clinical review of medical records and case histories that informs death certification in hospitals and other health-care settings. In addition, physician review has been the dominant approach because convincing alternatives have not been widely available. Most sites use more than one physician to review the VA responses and assign an underlying cause of death [8]. The use of physician review has served to mask the heterogeneity of VA instruments and content, as studies simply report the results of the physician assignment of causes of death. Validation studies have been published showing high sensitivity and specificity for selected causes [9–11], but even with high sensitivity and specificity, there are challenges in using physician-coded VA (PCVA) to derive unbiased estimates of population cause-specific mortality fractions [12–15].

Physician coding of VA data limits the potential use of VA for population heath monitoring in three important ways. First, local physician views on population epidemiology and causal hierarchies probably exert a profound influence on the diagnoses recorded. The influence of local medical culture is also seen in vital registration data with medical certification [16]. Strong views of local physicians will decrease the comparability of PCVA across populations. Differences in estimated cause-specific mortality rates may not be due to differences in the VA data themselves but simply because of differences in these local views. Second, the sensitivity and specificity of PCVA for deaths out of hospital may be much lower than currently recognized. Nearly all published validation studies have included collection of information from the household on medical records, death certificates, and recall of the cause of death by a health-care worker. Inclusion of this information in the VA will clearly exaggerate the performance of PCVA for deaths outside of hospital. Third, PCVA substantially increases the time and cost required to analyze VA data, limiting large-scale population applications.

There are three broad research directions underway to create more standardized underlying cause of death assignment using VA: creation of consensus standardized VA instruments; development of expert algorithms for assigning the underlying cause of death; and development of statistical methods. First, several initiatives are creating and testing standardized instruments for neonatal, child, and adult deaths [4,17]. Second, expert algorithms are based on the concept of distilling the process of physician review into standardized rules. These are meant to systematize the logic that an individual physician would use to assign a cause of death, but with the added advantage of consistent application [18,19].

The third approach is to develop statistical models to predict cause of death based on the detailed results of a standardized VA instrument [19–23]. A wide range of modeling strategies have been used, including logistic regression, neural networks, and Bayesian approaches. Byass et al. [21,24] have argued for the application of Bayes's theorem to individual cause-of-death assignments by using the symptom-level data recorded in the VA. Using data for rural Vietnam, that work demonstrated that such an approach has promise in categorizing individual causes of death when compared to PCVA. Recently, King and Lu [25] presented a sophisticated method for estimating cause-specific mortality fractions directly without individual cause-of-death attribution. Their method resolves the problem of generalizing VA analysis to the population based on test properties quantified in hospital validation studies. In this paper, we bring together the insights from King and Lu,

and Byass et al. into a single method, which we have termed the symptom pattern method (SP), to estimate both population-level cause-specific mortality fractions (CSMFs) and underlying causes of death for individuals. Using data from China, we investigate the performance of this method in comparison to PCVA at the population level and at the individual death level.

Methods

From a validation cause-of-death dataset for each variable included in the VA instrument (henceforth called symptoms), we can calculate the probability of household members responding in each response category for each cause of death. To simplify the explanation, let us assume all items in the VA instrument are dichotomous. Let S_{ij} be the probability that household members say yes to symptom i for cause of death j. If one considers symptom i like a diagnostic test, S_{ij} is the sensitivity of symptom i for disease j. Note that unlike most of the literature on VA, we are not referring to the sensitivity of the entire VA instrument subject to physician review, but rather to sensitivity of a single symptom in the instrument.

Let P_j be the CSMF for cause j, and S_1 be the proportion of all VA responses that say yes to symptom 1 in the population. Since the k causes of death are mutually exclusive and collectively exhaustive, the fraction of deaths that reported yes to symptom 1, S_1, must be the sum of the fraction of deaths that reported yes to symptom 1 and died of cause 1 (S_{11}), the fraction of deaths that reported yes to symptom 1 and died of cause 2 (S_{12}), and so on. This gives us

$$S_1 = S_{11}P_1 + S_{12}P_2 + ...S_{1k}P_k,$$

Since each death is in one and only one of the cause groups, it must also be true that

$$\sum_{j=1}^{k} P_j = 1,$$

and it then follows that

$$1 - \sum_{j=1}^{k-1} P_j = P_k.$$

Combining the relationships above, we see that for any set of $k - 1$ symptoms, the following set of linear equations must hold:

$$S_1 = S_{11}P_1 + S_{12}P_2 + ... + S_{1k}\left(1 - \sum_{j=1}^{k-1} P_j\right)$$

$$S_2 = S_{21}P_1 + S_{22}P_2 + ... + S_{2k}\left(1 - \sum_{j=1}^{k-1} P_j\right)$$

$$\vdots$$

$$S_{k-1} = S_{(k-1)1}P_1 + S_{(k-1)2}P_2 + ... + S_{(k-1)k}\left(1 - \sum_{j=1}^{k-1} P_j\right) \quad (1)$$

If the S_{ij} are known, this system has $k - 1$ equations and can be solved for the $k - 1$ unknown values of P_1 through P_{k-1}. If we have more than $k - 1$ symptoms, or in the presence of

measurement error, we could also estimate P_1 through P_{k-1} using regression techniques.

We wish to employ these equations to estimate the CSMFs, P_1 through P_{k-1}. A first pass at this problem might select a random subset of $k - 1$ symptoms and obtain values for P_j by estimating a regression, or by solving the system of equations. This process could be repeated, drawing many sets of $k - 1$ symptoms, resulting in many estimates of P_j. A final estimate for each P_j could then be obtained by computing the mean of the estimates gathered over all of the symptom draws. King and Lu advance this basic approach by randomly selecting sets of symptoms, called symptom profiles, and computing the S_i and S_{ij} values in the system of equations above for the symptom profiles, rather than for the individual symptoms [25]. This relaxes the assumption that, within a true cause of death group, the responses to individual symptoms are completely independent of one another. Over multiple draws of symptom profiles, then, King and Lu's strategy accounts for the correlation between groups of symptoms. Each random draw of symptom profiles yields an estimated set of P_j values. The mean of these P_j estimates over all of the iterations is then used as the best estimate of the CSMF for each cause j. We ran the algorithm with 300 draws of symptom profiles. We configured it to select 16 symptoms, each chosen from the set of symptoms with equal probability; the symptom profiles were then the combinations of those 16 symptoms that actually appeared in the data.

After obtaining CSMF estimates, we can use them as an input in a strategy to assign causes to individual deaths. For an individual death i, the probability of dying from cause j can be expressed following Bayes's theorem as:

$$P(D_i = j | S_i) = \frac{P(S_i | D_i = j)P(D_i = j)}{\sum_{j=1}^{k} P(S_i | D_i = j)P(D_i = j)}, \quad (2)$$

where $P(D_i = j | S_i)$ is the probability of individual i dying from cause j, conditional on the observed vector of symptom responses, S_i. In the terminology of Bayes's theorem, this is called the posterior. It can be calculated for each individual and cause of death. Since the causes of death are mutually exclusive and collectively exhaustive, the sum of these posterior probabilities for any individual death will be 100%. $P(S_i | D_i = j)$ is the probability of observing the vector of symptoms S_i conditional on individual i having died from cause j. If symptoms are largely based on the biology and natural history of the disease, this probability can be determined from well-designed studies in which a gold standard diagnosis is known. The expression $P(D_i = j)$, known as the prior, is the probability that individual i died of cause j, without taking into account the symptom response data. In this case, we use the CSMFs estimated with King and Lu's strategy, outlined in the discussion above, as the prior for individual-level cause of death assignment. In the same way that King and Lu use draws of symptom profiles to estimate the CSMFs, we repeatedly draw random combinations of symptoms, compute $P(S_i | D_i = j)$, and estimate posterior probabilities for each cause of death. We repeat this many times and take the mean of the results as our final estimate for the posterior probability for each cause. In the results we present here, we choose symptom profiles by selecting 15 symptoms from the entire set, with each symptom equally

Table 1. Verbal Autopsy Symptom Endorsement Rates by Cause-of-Death Group

Symptom	Hepatitis B	Tuberculosis	Lower Respiratory Infections	Other Group I[a]	Cerebrovascular Disease	Ischaemic Heart Disease	Diabetes Mellitus	Cirrhosis of Liver
Fever (%)	37	55	94	14	29	28	38	22
Breathless (%)	23	55	69	10	21	50	36	31
Breathless lying flat (%)	13	33	31	7	8	29	21	24
Palpitations (%)	23	39	25	3	20	62	39	31
Wheezing (%)	14	51	44	6	14	26	16	20
Cough (%)	10	88	75	3	11	28	28	16
Chest pain (%)	11	37	19	1	6	45	21	16
Diarrhea (%)	20	6	6	8	4	10	18	31
Bloody diarrhea (%)	16	4	0	0	2	2	4	10
Poor appetite (%)	49	49	38	8	25	30	39	59
Pain on swallowing (%)	0	2	13	0	3	4	6	6
Difficulty swallowing (%)	9	8	31	3	10	10	10	6
Headache (%)	16	14	0	6	31	12	11	8
Blood in urine (%)	9	6	6	6	4	1	6	6
Pain on urination (%)	3	8	6	0	2	3	9	8
Unable to urinate (%)	27	4	13	4	9	18	40	29
Urination too frequent (%)	9	10	19	1	11	15	28	22
Pins/needles on feet (%)	4	6	0	1	6	8	21	2
Abdominal pain (%)	33	4	6	4	3	5	6	37
Abdominal distension (%)	63	4	0	8	8	9	26	71
Vomiting (%)	19	8	13	7	26	20	28	25
Vomiting blood (%)	19	10	13	3	7	3	4	33
Abdominal mass (%)	6	0	0	0	1	2	0	2
Mentally confused (%)	16	16	0	1	11	6	5	10
Lost consciousness (%)	53	29	50	15	75	35	53	55
Paralyzed, one side (%)	1	2	13	3	44	11	14	2
Paralyzed, both legs (%)	1	0	6	3	13	5	8	2
Full body stiffness (%)	4	0	0	0	7	2	1	0
Neck pain (%)	4	4	0	0	8	6	4	2
Fits (%)	9	6	6	4	14	5	13	10
Ankle swelling (%)	37	39	44	8	21	27	44	41
Joint swelling (%)	7	10	13	0	5	9	21	16
Weight loss (%)	51	69	50	11	27	32	54	57
Mouth sores (%)	0	4	6	1	2	2	4	8
Looks pale (%)	24	24	13	10	16	21	34	25
Lip cyanosis (%)	10	14	38	4	16	33	19	25
Skin disease (%)	26	6	6	1	7	4	18	18
Face puffiness (%)	17	20	19	3	9	12	28	16
Both legs puffy (%)	44	39	19	7	17	29	49	49
Body puffiness (%)	13	8	0	1	3	7	23	14
Eyes yellow (%)	56	0	0	3	4	4	9	43
Experienced injury (%)	7	2	6	3	8	6	10	4
Chest pain detail (%)	1	12	0	1	3	30	8	6
Abdominal pain detail (%)	3	0	0	1	0	1	0	6
Abdominal distension detail (%)	3	0	0	0	2	1	1	10
Lost consciousness detail (%)	23	16	6	10	46	18	16	18
Ever a smoker (%)	49	73	56	10	42	44	30	41
Number in cause-of-death group	70	49	71	378	206	80	51	44

Each cell represents the endorsement rate for a given symptom and a given gold-standard cause of death.
[a]Groups I, II, and III refer to the Global Burden of Disease cause groupings. Group I is communicable, maternal, perinatal, and nutritional disorders; Group II is noncommunicable diseases; Group III is injuries [33].

likely to be chosen. The symptom profiles are then the observed combinations of those 15 symptoms that occur in the data. We draw symptom profiles 50 times. In order to uniquely assign an underlying cause of death to compute concordance rates and to compare to PCVA, for each individual, our cause-of-death estimate is then the cause that has the highest estimated posterior probability.

We compute population CSMFs two ways. First, we take the posterior probabilities estimated for each individual—that is, the output of the individual-level application of Bayes's theorem—and we aggregate these posterior probabilities by cause. We then use the relative distribution of posterior probabilities for each cause of death, across all of the individuals, as our updated estimate of the CSMFs. Second, we sum the number of deaths assigned to each cause based on which cause has the highest probability. The advantage of the first approach is that it more directly reflects the information content in the data about the presence of each cause in the population. The advantage of the second approach is that it is fully consistent with the assignment of a unique cause of

Table 1. Extended.

Symptom	Other Digestive Diseases	Colorectal Cancers	Liver Cancer	Stomach Cancer	Respiratory Cancers	Other Cancers	COPD	Other Respiratory Diseases
Fever (%)	39	38	29	32	37	37	35	45
Breathless (%)	39	32	32	21	58	38	77	55
Breathless lying flat (%)	34	13	22	6	38	28	49	29
Palpitations (%)	34	15	25	19	27	24	50	32
Wheezing (%)	27	13	20	9	38	29	65	32
Cough (%)	23	22	20	21	75	30	77	50
Chest pain (%)	9	13	13	10	35	25	19	11
Diarrhea (%)	20	40	20	15	6	18	8	3
Bloody diarrhea (%)	25	41	18	18	4	11	3	3
Poor appetite (%)	48	51	69	71	59	66	41	39
Pain on swallowing (%)	0	3	5	15	8	18	3	3
Difficulty swallowing (%)	9	7	8	25	16	30	11	16
Headache (%)	11	10	12	12	16	14	14	8
Blood in urine (%)	5	9	5	3	3	10	4	3
Pain on urination (%)	2	1	3	1	1	7	7	5
Unable to urinate (%)	20	16	26	26	11	19	18	16
Urination too frequent (%)	16	9	15	9	7	11	16	8
Pins/needles on feet (%)	5	7	6	6	6	1	4	3
Abdominal pain (%)	41	51	47	50	8	25	4	5
Abdominal distension (%)	25	60	67	43	17	30	15	11
Vomiting (%)	39	41	34	41	19	31	8	5
Vomiting blood (%)	14	10	30	21	8	11	2	5
Abdominal mass (%)	2	41	35	35	8	18	1	0
Mentally confused (%)	9	15	9	6	8	11	10	13
Lost consciousness (%)	45	29	53	51	30	37	36	26
Paralyzed, one side (%)	7	3	2	1	8	7	8	16
Paralyzed, both legs (%)	0	4	0	0	3	3	4	5
Full body stiffness (%)	0	1	4	1	2	1	3	5
Neck pain (%)	0	1	3	3	8	6	3	8
Fits (%)	5	4	5	3	4	7	5	13
Ankle swelling (%)	45	46	48	44	32	42	44	29
Joint swelling (%)	9	12	16	6	6	10	11	8
Weight loss (%)	55	75	75	87	73	70	50	39
Mouth sores (%)	2	0	4	1	2	7	6	8
Looks pale (%)	36	35	26	50	32	41	18	24
Lip cyanosis (%)	27	7	25	7	27	17	41	34
Skin disease (%)	11	10	24	7	13	21	8	18
Face puffiness (%)	18	10	14	12	12	15	26	13
Both legs puffy (%)	41	54	47	40	33	42	39	32
Body puffiness (%)	9	12	14	16	9	7	10	8
Eyes yellow (%)	11	25	52	21	8	27	6	3
Experienced injury (%)	9	7	2	6	4	2	7	8
Chest pain detail (%)	0	3	6	4	9	4	5	0
Abdominal pain detail (%)	9	4	1	16	1	4	0	0
Abdominal distension detail (%)	14	10	7	6	3	2	3	5
Lost consciousness detail (%)	20	9	25	10	9	11	15	13
Ever a smoker (%)	41	37	49	51	70	41	61	34
Number in cause-of-death group	68	102	68	212	138	184	101	81

death at the individual level. We present results from the first strategy and evidence that shows that the two strategies yield very similar results.

To illustrate this approach, we have used data from our China VA validation study based on medical record review [11]. For the 2,089 deaths with a gold standard underlying cause of death, and a completed VA instrument, we have divided deaths into 23 mutually exclusive and collectively exhaustive cause groups (see Table 1). These groups have been selected such that there were at least 15 deaths in each cause group, including the combined 'other' groups. Based on the VA instrument, 47 dichotomous symptom variables have

been selected. We have first tested the validity of our SP method by sampling with replacement the 2,089 deaths to generate two datasets: a validation dataset of 4,600 deaths, 200 for each cause, to estimate the values of S_{ij}, and a community dataset of 10,000 deaths, to which we apply the SP method. One limitation of our use of simulation datasets is that single deaths from the original dataset may be sampled multiple times, resulting in less variation in symptom profiles than we would expect to see in real data.

The gold standard underlying cause of death was derived through expert review of the medical records by a panel of three physicians. Medical records were abstracted by trained

Table 1. Extended.

Symptom	Nephritis	Endocrine Disorders	Hypertensive Heart Disease	Rheumatic Heart Disease	Acute Pancreatitis	Other Group II[a]	Injuries—All Group III[a]	All Deaths
Fever (%)	33	50	36	33	31	28	21	33
Breathless (%)	27	40	45	76	19	38	14	39
Breathless lying flat (%)	24	20	29	43	25	19	6	23
Palpitations (%)	27	50	57	67	25	28	14	31
Wheezing (%)	30	15	29	19	13	18	14	26
Cough (%)	30	15	31	38	25	18	9	32
Chest pain (%)	12	10	31	14	6	14	4	19
Diarrhea (%)	12	15	21	5	13	12	2	12
Bloody diarrhea (%)	6	15	10	0	0	8	0	7
Poor appetite (%)	39	40	50	29	31	43	21	42
Pain on swallowing (%)	0	10	2	0	0	3	2	5
Difficulty swallowing (%)	3	5	14	5	0	14	5	12
Headache (%)	12	25	24	14	6	14	14	16
Blood in urine (%)	12	10	12	0	6	8	2	5
Pain on urination (%)	15	10	10	0	0	5	1	4
Unable to urinate (%)	52	30	43	29	25	21	9	18
Urination too frequent (%)	33	20	19	5	0	8	7	12
Pins/needles on feet (%)	12	10	12	0	0	7	1	6
Abdominal pain (%)	9	5	12	10	63	21	1	15
Abdominal distension (%)	21	15	29	19	56	16	5	22
Vomiting (%)	27	30	24	19	50	25	10	23
Vomiting blood (%)	0	15	7	0	6	5	5	9
Abdominal mass (%)	0	0	0	0	13	2	4	7
Mentally confused (%)	9	15	14	14	6	9	11	9
Lost consciousness (%)	58	50	52	29	31	39	62	46
Paralyzed, one side (%)	9	10	12	14	6	9	10	14
Paralyzed, both legs (%)	3	0	2	14	6	3	5	5
Full body stiffness (%)	6	0	5	0	0	7	10	3
Neck pain (%)	3	5	10	5	0	3	2	5
Fits (%)	9	5	10	10	13	11	7	8
Ankle swelling (%)	45	45	60	62	38	25	20	34
Joint swelling (%)	21	10	19	19	6	8	12	9
Weight loss (%)	48	50	55	48	44	42	17	49
Mouth sores (%)	9	15	7	0	0	3	0	3
Looks pale (%)	39	50	45	24	31	26	16	26
Lip cyanosis (%)	15	20	33	43	19	19	9	22
Skin disease (%)	30	40	10	5	19	12	15	12
Face puffiness (%)	36	5	50	14	13	16	11	15
Both legs puffy (%)	61	25	62	57	25	25	14	33
Body puffiness (%)	18	0	29	0	19	11	7	9
Eyes yellow (%)	3	35	10	0	13	7	2	13
Experienced injury (%)	9	0	10	5	0	6	73	8
Chest pain detail (%)	0	5	10	14	0	2	1	7
Abdominal pain detail (%)	0	0	5	0	13	3	0	2
Abdominal distension detail (%)	0	5	17	5	25	6	0	4
Lost consciousness detail (%)	21	10	19	14	0	21	42	22
Ever a smoker (%)	33	30	43	14	6	30	37	45
Number in cause-of-death group	16	33	38	20	42	21	16	2,089

personnel including clinical information, diagnostic results, and treatment details. Expert review death certificates were coded by the World Health Organization (WHO) Collaborating Centre for Classification of Diseases in Beijing to select and code the underlying cause of death according to the International Classification of Diseases, Tenth Revision (ICD-10) [11,26]. Further information is available in Rao et al. [27].

As in most VA studies that use PCVA, the instrument used in China also collected information on medical record recall from the household. This information includes items on cause of death from the death certificate and burial permit, hospital forms, and the child's health card. Information from at least one of these sources was available for 94% percent of

the deaths. We have repeated our analysis using both the full VA data and the medical record history data by treating the recall of medical records as if they were simply more S_{ij} values. To provide a comparison to PCVA, two Chinese physicians recoded the 2,089 VAs blinded to medical record recall. This empirical work also provides a test of the importance of medical record recall on PCVA as opposed to information contained in the symptom recall.

To further test the sensitivity of the SP method to the cause composition of mortality in a community, we have used sampling with replacement of the 2,089 China validation dataset deaths with weights by cause to alter the background cause composition for the community sample. We have used

Table 2. Verbal Autopsy Medical Record Recall (MRR) Symptom Endorsement Rates by Cause-of-Death Group

Symptom	Hepatitis B	Tuberculosis	Lower Respiratory Infections	Other Group I[a]	Cerebrovascular Disease	Ischemic Heart Disease	Diabetes Mellitus	Cirrhosis of Liver
Respiratory failure (%)	6	12	44	3	18	15	20	4
Circulatory failure (%)	10	8	13	7	15	27	20	4
Lung cancer (%)	0	4	0	0	0	0	0	0
Brain hematoma (%)	0	0	0	0	78	10	13	6
Diabetes (%)	0	0	0	0	6	10	65	4
Angina/heart attack (%)	1	2	6	0	6	65	6	2
Pneumonia (%)	1	12	75	3	8	7	13	2
Kidney disease (%)	4	0	0	1	4	8	44	12
Multiple organ failure (%)	6	0	19	3	4	3	4	8
Heart disease (%)	0	14	0	3	3	12	3	0
Hypertension (%)	3	2	0	0	15	6	5	0
Liver cancer (%)	3	0	0	0	0	0	0	0
Liver cirrhosis (%)	50	0	0	7	1	0	1	63
COPD/asthma/bronchitis (%)	0	14	13	0	2	7	3	0
Stomach cancer (%)	0	0	0	0	0	1	1	2
Hemorrhage (%)	11	4	0	0	2	0	1	24
Hepatitis B (%)	44	0	0	1	0	0	1	12
Ascites (%)	19	0	0	0	0	0	0	20
Tuberculosis (%)	0	63	0	0	0	1	0	0
Stomach ulcer (%)	1	0	0	0	0	0	0	6
Cancer of esophagus (%)	0	0	0	0	0	0	0	0
Liver failure (%)	17	0	0	4	0	0	0	14
Rectal cancer (%)	0	0	0	0	0	0	0	0
Aplastic anemia (%)	0	0	0	0	0	0	0	0
Colon cancer (%)	0	0	0	0	0	0	0	0
Breast cancer (%)	0	2	0	0	0	0	0	0
Pancreatic head carcinoma (%)	0	0	0	0	0	0	0	0
Tumors (%)	1	0	0	0	2	0	0	0
Septic shock (%)	1	0	0	1	0	0	3	0
Fever/infection (%)	1	12	0	7	2	0	1	2
Injury (%)	4	4	6	7	8	1	0	6
Kidney necrosis (%)	0	0	0	0	1	0	1	0
Pancreatic necrosis (%)	1	0	0	0	0	0	0	0
Rheumatic heart disease (%)	0	0	0	0	0	0	0	0
Hyperthyroidism (%)	0	0	0	0	0	0	0	0
Number in cause-of-death group	70	49	71	378	206	80	51	44

Each cell represents the endorsement rate for a given symptom and a given gold-standard cause of death. The symptoms were derived from recall of medical records and respondent interactions with the health system.
*Groups I, II, and III refer to the Global Burden of Disease cause groupings. Group I is communicable, maternal, perinatal, and nutritional disorders; Group II is noncommunicable diseases; Group III is injuries [33].
doi:10.1371/journal.pmed.0040327.t002

weights based on the WHO estimates of cause of death composition for South East Asia subregion D (SEAR-D; primarily India) and Western Pacific subregion B (WPR-B; primarily China). For the same community samples, we have calculated the same results for PCVA as well as for the SP method.

Three metrics of predictive validity have been computed: for the population-level results, we calculate the average percent relative error and the average absolute error for the 23 CSMFs. The average percent relative error is the average amount that the estimated CSMF deviates from the true CSMF, as a fraction of the true CSMF. In other words, it is the average, over all of the causes, of the quantity

$$\left| \frac{CSMF_{est} - CSMF_{true}}{CSMF_{true}} \right|. \qquad (3)$$

The average absolute error is the average absolute amount that the estimated CSMF deviates from the true CSMF, over all of the causes of death. In other words, it is the average, over all of the causes, of the quantity

$$|CSMF_{est} - CSMF_{true}|. \qquad (4)$$

Performance of the individual cause of death assignment has been evaluated using the concordance rate: the percentage of times that the true cause, based on the hospital record review, has been correctly assigned. Because this step produces a probability distribution across causes, we calculate the concordance rate based on the cause of death assigned the highest probability of death.

Results

Table 1 provides the S_{ij} values from the China VA study for the 47 main symptoms. We have excluded from this analysis

Table 2. Extended.

Symptom	Other Digestive Diseases	Colorectal Cancers	Liver Cancer	Stomach Cancer	Respiratory Cancers	Other Cancers	COPD	Other Respiratory Diseases
Respiratory failure (%)	9	15	11	15	24	16	42	45
Circulatory failure (%)	11	12	10	7	10	19	39	24
Lung cancer (%)	2	4	8	1	83	3	3	0
Brain hematoma (%)	5	1	1	0	1	1	4	5
Diabetes (%)	5	3	1	0	2	0	5	5
Angina/heart attack (%)	5	0	3	0	3	1	5	5
Pneumonia (%)	11	1	0	0	4	5	18	42
Kidney disease (%)	11	3	4	4	1	3	7	8
Multiple organ failure (%)	11	10	12	12	7	7	5	5
Heart disease (%)	7	0	3	1	1	1	25	11
Hypertension (%)	5	0	3	0	1	1	2	0
Liver cancer (%)	0	10	77	1	2	4	0	0
Liver cirrhosis (%)	2	1	11	1	0	1	0	0
COPD/asthma/bronchitis (%)	2	0	4	1	3	2	54	29
Stomach cancer (%)	0	1	0	76	0	1	0	0
Hemorrhage (%)	23	0	8	1	1	3	1	0
Hepatitis B (%)	2	0	4	0	0	1	0	0
Ascites (%)	0	4	4	3	0	0	1	0
Tuberculosis (%)	0	0	0	0	0	0	2	0
Stomach ulcer (%)	0	0	0	1	0	1	0	0
Cancer of esophagus (%)	0	0	0	4	0	17	1	0
Liver failure (%)	5	0	5	0	0	1	0	0
Rectal cancer (%)	0	35	0	0	0	0	0	0
Aplastic anemia (%)	0	1	0	0	0	1	0	0
Colon cancer (%)	0	29	0	0	0	0	0	0
Breast cancer (%)	0	0	0	0	0	10	0	0
Pancreatic head carcinoma (%)	0	1	0	0	0	2	0	0
Tumors (%)	5	22	5	13	6	51	0	0
Septic shock (%)	7	1	1	0	0	0	0	3
Fever/infection (%)	9	0	1	0	0	1	2	0
Injury (%)	2	1	1	0	0	1	1	3
Kidney necrosis (%)	0	0	0	0	0	0	0	0
Pancreatic necrosis (%)	2	0	0	0	0	1	0	0
Rheumatic heart disease (%)	0	0	0	0	0	0	0	0
Hyperthyroidism (%)	0	1	0	0	0	0	0	0
Number in cause-of-death group	68	102	68	212	138	184	101	81

information on the duration of symptoms. We undertook extensive analysis using symptom duration data as further S_{ij} values, but this did not improve the performance of the method over and above inclusion of just the symptoms. We can think of each entry in the table as the properties of the symptom (row) for the cause of death (column). For example, the entry at the fifth row and third column is 44%, meaning that 44% of VA respondents for deaths from lower respiratory infections responded yes when asked if the deceased wheezed. There is clearly considerable information content in the individual symptoms even without information on the sequencing, duration, and clustering of symptoms. There is also clear evidence that, like any survey instrument, there is a significant background noise for each symptom. Injuries, whose occurrence can be expected to be uncorrelated with many symptoms, provide insight into these background rates. For example, among deaths due to injuries, 21% of household members reported fever, 21% poor appetite, 17% weight loss, and 15% skin disease; only 73% reported that the decedent suffered an injury. The latter surprising finding has also been observed in South Africa [28]. Some symptoms, which medical knowledge would suggest are highly specific such as paralysis on one side, are present for more than 10% of decedents for seven of 27 causes, suggesting that this symptom has a different cultural interpretation in China.

Table 2 summarizes the information from the last part of the VA instrument on the recall of causes of death in the death certificate, burial certificate, and household recall. We have combined the data from all these sources into a single measure, as they all reflect information from providers or government registries as reflected through the recall and experience of household members. These symptoms are our coding of the information that physicians with access to medical record recall had; each symptom is a phrase that appears in the medical records, on the death certificate, or on the burial certificate. We identified key phrases used in the medical records, death certificates, and burial certificates and mapped these phrases to ICD categories. It is possible that meaningful local phrases may have been missed in this mapping, which would tend to bias our analysis against the SP method. The physicians who were blinded to medical record recall did not have access to this information, and our results without medical record recall do not make use of it. These

Table 2. Extended.

Symptom	Nephritis	Endocrine Disorders	Hypertensive Heart Disease	Rheumatic Heart Dis.	Acute Pancreatitis	Other Group II[a]	Injuries, All Group III[a]	All Deaths
Respiratory failure (%)	9	5	14	14	13	14	19	18
Circulatory failure (%)	18	10	33	43	13	16	10	17
Lung cancer (%)	0	0	0	5	0	1	0	10
Brain hematoma (%)	0	10	17	24	0	6	23	18
Diabetes (%)	0	5	14	5	6	0	5	6
Angina/heart attack (%)	6	10	19	5	0	3	1	10
Pneumonia (%)	6	5	10	14	6	7	6	8
Kidney disease (%)	73	10	52	5	25	19	5	9
Multiple organ failure (%)	0	10	5	5	13	7	5	6
Heart disease (%)	3	0	10	10	6	6	0	6
Hypertension (%)	3	0	26	14	0	3	1	5
Liver cancer (%)	0	0	0	0	0	0	0	5
Liver cirrhosis (%)	0	0	0	0	6	0	1	4
COPD/asthma/bronchitis (%)	3	0	2	5	0	4	0	8
Stomach cancer (%)	0	0	0	0	0	1	1	3
Hemorrhage (%)	0	5	0	0	0	6	0	3
Hepatitis B (%)	0	5	0	0	0	0	0	2
Ascites (%)	0	0	0	0	0	1	1	2
Tuberculosis (%)	0	0	0	0	0	0	1	2
Stomach ulcer (%)	0	0	0	0	0	2	0	0
Cancer of esophagus (%)	0	0	0	0	0	0	0	1
Liver failure (%)	0	0	0	0	6	0	0	2
Rectal cancer (%)	0	0	0	0	0	0	0	1
Aplastic anemia (%)	0	60	0	0	0	0	0	1
Colon cancer (%)	0	0	0	0	0	0	0	1
Breast cancer (%)	0	0	0	0	0	0	0	1
Pancreatic head carcinoma (%)	0	0	0	0	0	0	0	0
Tumors (%)	3	0	2	5	0	4	1	6
Septic shock (%)	0	0	0	0	0	0	1	1
Fever/infection (%)	0	0	5	0	0	4	0	2
Injury (%)	0	0	2	0	0	5	54	5
Kidney necrosis (%)	12	0	0	0	6	1	0	0
Pancreatic necrosis (%)	0	0	0	0	69	0	0	1
Rheumatic heart disease (%)	0	0	0	43	0	0	0	1
Hyperthyroidism (%)	0	0	0	0	0	0	1	0
Number in cause-of-death group	16	33	38	20	42	21	16	2,089

can be interpreted in the same way as Table 1. So, an endorsement rate of 4% for the third row (lung cancer) and second column (TB) means that for 4% of deaths whose underlying cause was TB, the medical records, burial certificate, or death certificate mentioned lung cancer.

Figure 1 summarizes the application of the SP method and PCVA to a 10,000 community sample of deaths in China. With medical record recall, the SP method estimates the true CSMFs with average relative error of about 16%, while PCVA gives relative error of about 27%. In absolute terms, the average difference between the estimated and the true CSMF is 0.7% for the SP method, and 1.1% for PCVA. At the individual level, PCVA correctly assigns the cause of death in slightly more than two-thirds of the cases, while the SP method is correct for over 80% of the deaths. The dependence of the physician assignment of cause of death based on the recall of medical records is evident in the dramatic change in PCVA performance when the medical records are excluded. The average relative error for the CSMFs increases to over 70% and the concordance rate at the individual death level drops to below 40%. The SP method estimates of the

CSMFs is unaffected by the exclusion of the medical record recall; the relative error remains at about the same level, 14%, and the absolute error, at 0.6%, is similar as well. The individual concordance rate drops slightly, to just under 80% which is still significantly better than PCVA. But the reality is that symptoms alone are probably not a sufficient basis for assigning a unique cause of death to each individual; diagnostic tests and imaging studies are critical for diagnosis in many cases.

A long-standing concern with VA validation studies has been that sensitivity and specificity of PCVA may vary as a function of the cause composition of mortality [12,29]. To explore this problem, we have varied the background cause composition of the community sample of deaths generated through sampling with replacement with cause of death composition weights that match the SEAR-D and WPR-B WHO subregions. Figures 2 and 3 summarize the three measures of predictive validity: CSMF average relative error, CSMF average absolute error, and individual concordance rates. The results for both regions demonstrate that the SP method produces quite accurate estimates of CSMFs and

(A)

(A)

(B)

(B)

(C)

(C)
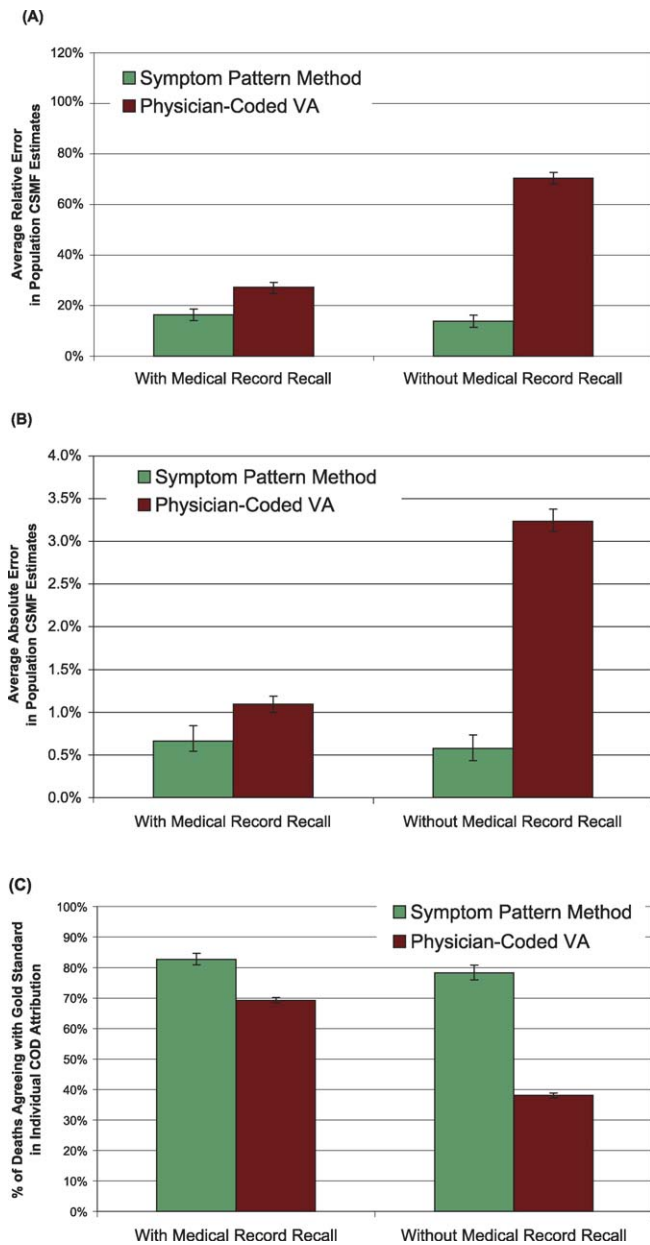
Figure 1. Summary Measures of Error in Population and Individual-Level Cause-of-Death Estimates for China Data
Results are from the SP method and PCVA with and without the symptoms derived from recall of medical records.
(A) Average relative error in population-level CSMF estimates for China data.
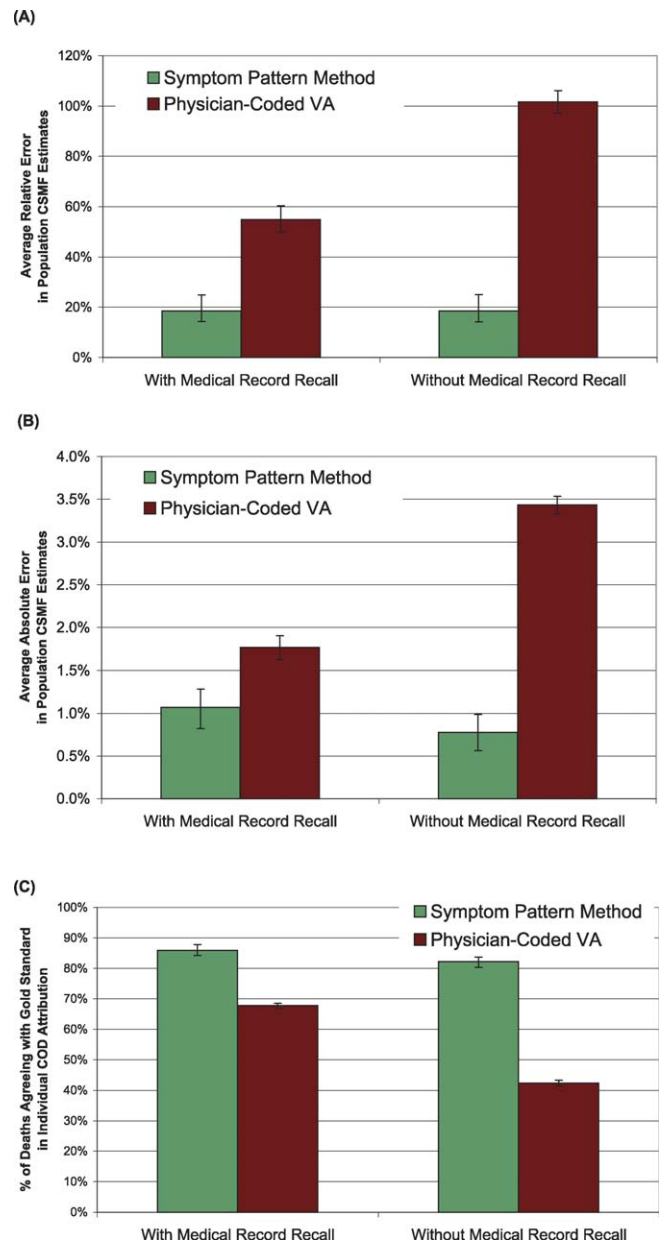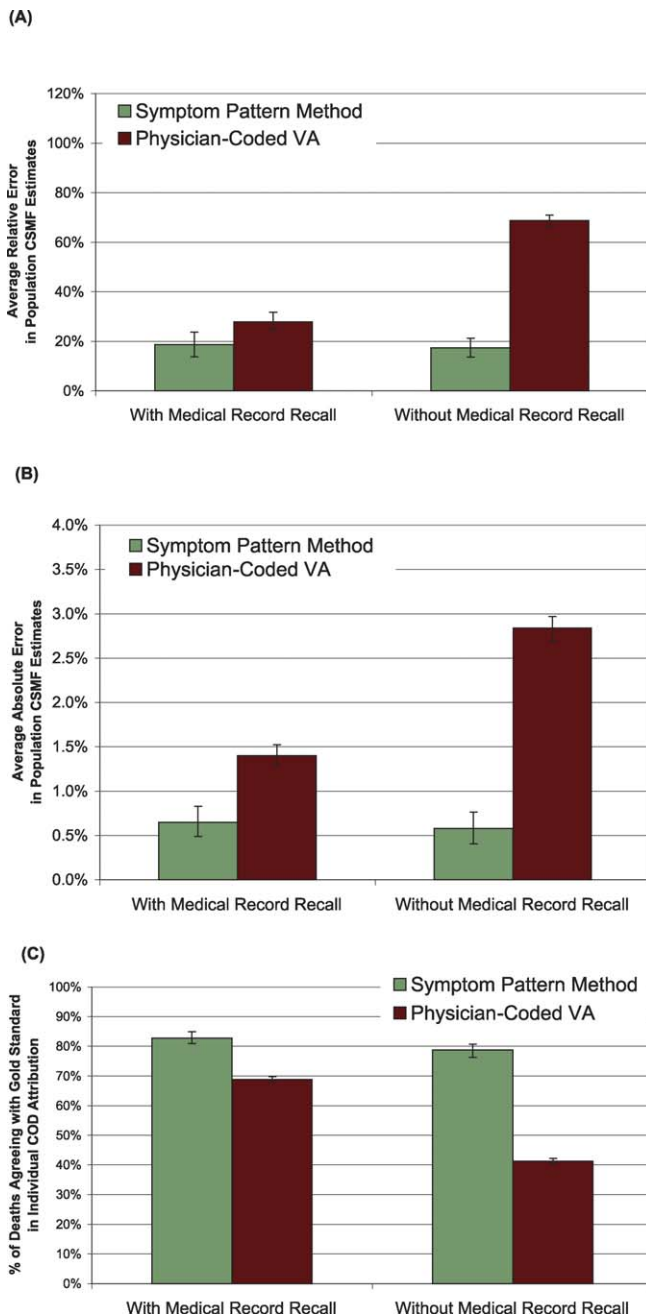(B) Average absolute error in population-level CSMF estimates for China data.
(C) Percent concordance (percent of deaths agreeing with gold standard) for individual-level cause-of-death (COD) assignments in China data.
doi:10.1371/journal.pmed.0040327.g001

Figure 2. Summary Measures of Error in Population and Individual-Level Cause-of-Death Estimates, Background Cause Composition of Mortality Changed to Simulate Distribution in WHO Subregion SEAR-D
Results are from the SP method and PCVA, with and without the symptoms derived from medical record recall.
(A) Average relative error in population-level CSMF estimates for SEAR-D.
(B) Average absolute error in population-level CSMF estimates for simulated SEAR-D data.
(C) Percent concordance (percent of deaths agreeing with gold standard) for individual-level cause of death (COD) assignments in SEAR-D data.
doi:10.1371/journal.pmed.0040327.g002

individual concordance rates that are over 75% with no medical record recall and over 80% with medical record recall. At least based on these two simulation studies, there is no evidence that the SP method is sensitive to changing the cause composition of mortality.

Table 3 compares individual concordance rates for the China community simulated sample by cause of death with and without medical records for PCVA and the SP method. For nine of the causes of death, the concordance rates for

physicians drops by more than 75% including: hepatitis B, diabetes, colorectal cancer, stomach cancer, lower respiratory infections, hypertensive heart disease, endocrine disorders, rheumatic heart disease, and acute pancreatitis. For some causes, such as hepatitis B and acute pancreatitis, PCVA correctly assigns no deaths when medical record recall is removed. For the SP method, the lowest concordances are in other group I conditions, at 34%, other group II conditions, at 61%, and lower respiratory infections, at 65%. We would

Figure 3. Summary Measures of Error in Population and Individual-Level Cause-of-Death Estimates, Background Cause Composition of Mortality Changed to Simulate Distribution in WHO Subregion WPR-B

Results are from the SP method and PCVA, with and without the symptoms derived from medical record recall.
(A) Average relative error in population-level CSMF estimates for WPR-B.
(B) Average absolute error in population-level CSMF estimates for simulated WPR-B data.
(C) Percent concordance (percent of deaths agreeing with gold standard) for individual-level cause of death assignments in WPR-B data.
doi:10.1371/journal.pmed.0040327.g003

generally expect SP to perform least well on residual categories of death like other group I conditions. These residual categories represent an average symptom profile of distinct diseases included in the residual. Much of the distinct information content in the set of S_{ij} values is thus lost due to averaging across different entities. The poor results for lower

respiratory infections may be due to the small number of cases, 16, in the original validation dataset. Without medical record recall, the same three conditions perform worst, at 53%, 64%, and 70% respectively.

Figure 4 gives an example of the posterior distributions produced by SP for two deaths in one of the simulated datasets. The gold-standard cause of death for both of these is cirrhosis of the liver. Since we assigned the most probable single cause to each death, we can see that SP would correctly assign the cause in the first death, where cirrhosis is the most probable posterior cause, while SP would be incorrect for the second death, where cerebrovascular disease is most probable in the posterior. The distribution for the second death is spread across the causes more evenly, meaning that the symptoms were not as predictive as they were for the first death, where they are more focused on cirrhosis, hepatitis B, and liver cancer. Although we have only used the most probable posterior cause in assigning a single cause to each death, these posterior distributions have information that could be exploited for other purposes. For example, the diffuseness of the distribution gives us some idea of the certainty of the cause assignment.

As noted in the Methods section, we can calculate population CSMFs by summing the number of deaths assigned to each cause based on the highest probability cause of death in the posterior distribution or we can sum the posterior probabilities for each death across each cause. Figure 5 compares these two approaches for the 23 causes in the China data. For each cause, it shows the average CSMF estimated for the 100 draws of data obtained by using the single highest probability cause to the CSMF obtained by using the full posterior distribution. The results are very similar, with a correlation of 0.98. The most noticeable outlier is the CSMF for cerebrovascular disease, which is generally estimated to be larger when the single most probable cause is used.

Discussion

Overall, the SP method does remarkably well. At the population level, it provides more accurate estimates of cause-specific mortality fractions than PCVA; indeed, at least in China, PCVA is not a viable method for estimating the cause composition of mortality at the population level. At the individual level, it clearly outperforms PCVA in assigning causes of death to individual responses, using the same information. The difference is most striking when the two strategies are not allowed access to information from household recall of medical records, which is the test most relevant to implementing verbal autopsy in a resource-poor setting. Based on this analysis of data from China, the superiority of the SP method for estimating CSMFs and assigning causes of death at the individual level does not appear to be affected by changing the background cause of death composition in the population sample. In contrast, in China PCVA appears to be extremely sensitive to the availability of household recall of medical records.

Cause of death assignment for some deaths must be highly uncertain, because the information content in signs and symptoms is limited. While we have emphasized in presenting results the comparison of PCVA with SP using the highest-probability cause for each death, this method generates an

Table 3. Individual Concordance (Percent Agreement with Gold Standard Cause of Death), by Gold Standard Cause of Death in China Community Simulated Sample

Cause of Death	Symptom Pattern Method (%)		Physician-Coded VA (%)	
	With MRR	Without MRR	With MRR	Without MRR
Hepatitis B	91	86	37	0
Tuberculosis	88	79	67	22
Lower respiratory infections	65	70	73	12
Other Group I	34	53	70	66
Cerebrovascular disease	88	84	81	66
Ischaemic heart disease	79	71	62	35
Diabetes mellitus	86	83	56	2
Cirrhosis of liver	94	92	75	26
Other digestive diseases	85	91	47	31
Colorectal cancers	91	89	74	15
Liver cancer	90	84	86	27
Stomach cancer	95	84	91	19
Respiratory cancers	88	72	89	45
Other cancers	85	76	83	59
COPD	80	74	50	33
Other respiratory diseases	70	76	35	18
Nephritis	84	84	59	28
Endocrine disorders	82	83	54	8
Hypertensive heart disease	93	94	38	0
Rheumatic heart disease	72	75	57	3
Acute pancreatitis	88	78	46	0
Other Group II	61	64	48	21
Injuries—all Group III	84	84	80	74
All Deaths	83	78	69	38

Results are from the SP method and PCVA, with and without the symptoms derived from recall of medical records (MRR).
doi:10.1371/journal.pmed.0040327.t003

uncertainty distribution across causes, as illustrated for two deaths in Figure 4. These distributions are a direct reflection of the information content in the VA responses. Future studies may be able to make direct use of this information to guide a deeper recognition of which causes can be more accurately identified on the basis of signs and symptoms alone. Conventions may be developed such that deaths for which the uncertainty distribution is diffuse across many causes would be assigned to an unknown or ill-defined category. In any case, the SP method provides a more realistic reflection of the information content in the VA dataset.

Reanalysis of the China VA validation data by physicians blinded to medical record recall shows that physicians are highly influenced by this information, as one might expect. Most published validation studies on PCVA include medical record recall. The levels of sensitivity, specificity, or overall concordance rates reported in the literature for deaths with medical record recall are substantially overstated for deaths that occur without contact with medical services. Some authors [5] have even claimed that high inter-rater reliability between physicians is evidence of the validity of PCVA. Of course, high inter-rater reliability could simply be due to multiple physicians using the medical record recall to assign the cause of death, not the symptom patterns. As such, inter-rater reliability provides no direct evidence of validity.

Because the SP method does not use physicians in assigning causes of death, it has the potential to dramatically reduce the cost of VA at the population level and to increase the comparability of results across populations. Use of VA in routine public health surveillance has in part been limited by

the availability of standardized international instruments but more so by the practicality of using physicians trained in the review of VA data. The SP method requires no individual coders and will reduce cost and implementation issues. Another major advantage is that it can be implemented in a completely replicable manner in different populations. By standardizing how causes of death are assigned and removing the unquantifiable influence of physician prior beliefs from the results, comparability across populations within a country and across countries will be substantially enhanced.

The critical requirement for our method is that we know the set of symptom level sensitivities S_{ij} values or that we can adequately predict these sensitivities for a population. This is an important assumption with several implications. First, we need to measure the S_{ij} values in a number of populations and to measure them for as many detailed causes as possible. More data on stroke S_{ij} values are not likely to be very useful, but more for diseases such as hepatitis B or COPD would be. Measuring the S_{ij} values may also be challenging in resource-poor settings where deaths coded to a reliable gold standard are difficult to locate using existing hospital medical records. To address this issue ongoing field studies, in Tanzania, India, and the Philippines, are defining more carefully strict diagnostic criteria for gold standard cause attribution. Second, we need to study the variability of the S_{ij} values as a function of culture, individual respondent characteristics such as income, education or medical knowledge, and decedent characteristics such as the experience of health care. Large validation datasets such as the one being collected in the Grand Challenges in Global Health 13 project [30] may
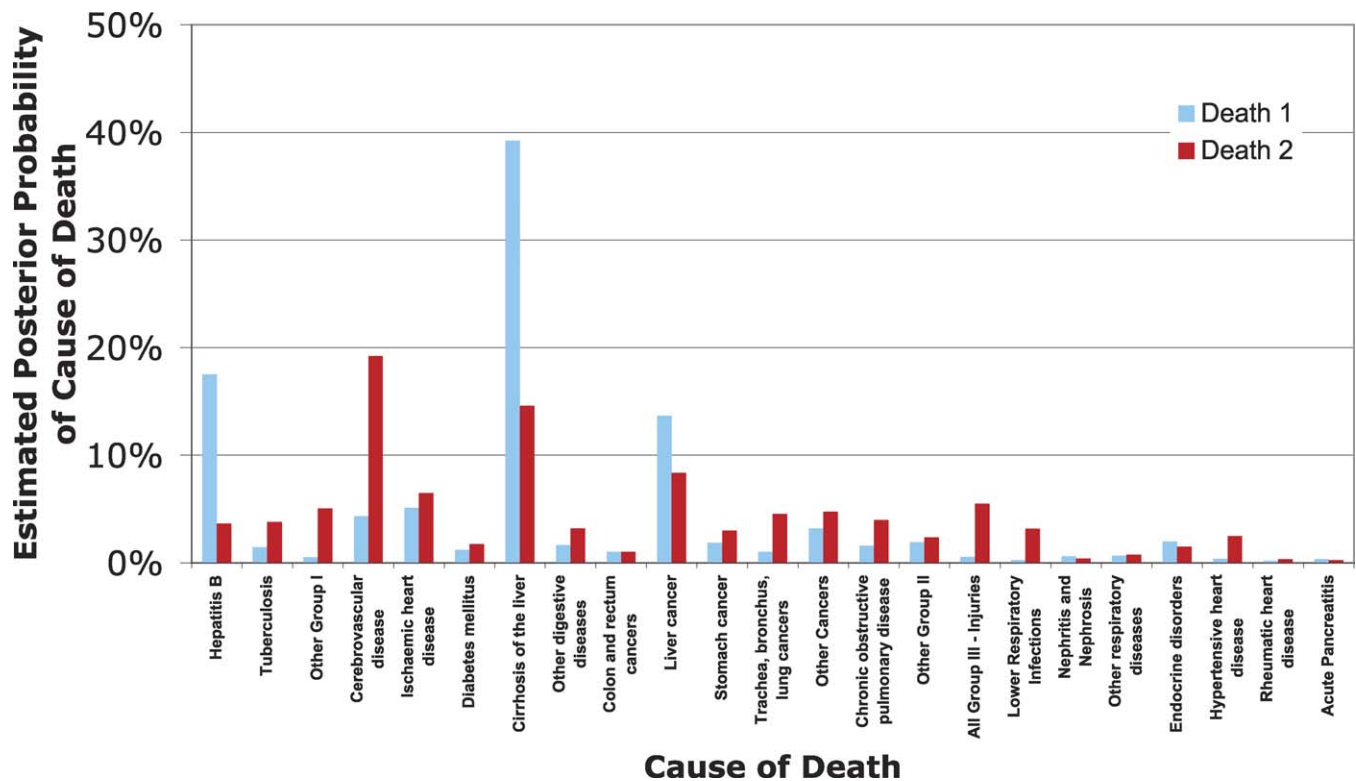
Figure 4. Sample Posterior Cause-of-Death Probabilities Produced by the Symptom Pattern Method for Two Deaths Whose True Cause Was Cirrhosis of the Liver

The height of each bar over each cause of death represents the probability, estimated using the SP method, of the individual having died of that cause.
doi:10.1371/journal.pmed.0040327.g004

provide an opportunity to study the predictors of the S_{ij} values. Third, symptom responses may be profoundly affected by disease natural history but they may also potentially be affected by cultural factors; that is, it may be that the items on the VA questionnaire will behave differently as a result of culture. In these situations, local validation studies will be needed to define how a VA item functions in a given population.

New data are needed in different populations to confirm the success of the SP method. In this study, we simulated data for the Chinese population and two WHO subregions based on a sample of VA responses from China; further data collection will allow us to test the generalizability of these results to other parts of the world.

We suggest that VA validation studies need to be radically redesigned. Most importantly, samples need to be chosen so as to get approximately equal numbers for important causes of death rather than a random sample of the population if we want to accurately characterize S_{ij}. This is an important reversal of the usual validation study in which a sequential sample of deaths in hospital is often used. As noted, stricter review processes to determine the gold standard diagnosis are also needed, so that only cases with medical records in which the underlying cause can be determined with substantial certainty are included. Such stricter gold standards are more feasible to implement in low-resource settings when the number of deaths meeting the gold standard for each cause can be optimized.

In most cases, validation studies are undertaken for deaths

that have occurred in hospital, and household recall of the hospital experience is included in what is presented to the physicians. The key exception to this has been validation studies of VA that seek to categorize deaths into only two causes, AIDS and non-AIDS, where HIV serostatus in a cohort is used as a the gold standard [27,31,32]. When this information is excluded from the analysis, our CSMFs are still as accurate but individual cause of death concordance drops. In other words, VA at the individual level may not work as well as published validation studies suggest when applied to deaths of individuals who did not have contact with health services prior to death. This is not a limitation of PCVA or of our approach. Rather, it may simply be a reflection of the true information content in the symptom recall. Further research studies, particularly on PCVA excluding all hospital and death record information, will help to better understand this effect.

Our approach to VA cause of death assignment should be tested and validated in other populations outside of China. This will be greatly facilitated if research teams are encouraged to put the full microdata from validation studies into the public domain. We also see a need for a multi-country validation exercise and a series of cause-of-death surveys to derive S_{ij} data, possibly using the newly developed WHO standardized tools. If our methods are confirmed in other populations, the cost of obtaining useful, comparable cause of death information in poor countries will be greatly reduced. Combined with advances being made on verbal autopsy instrument design, this research has the potential to
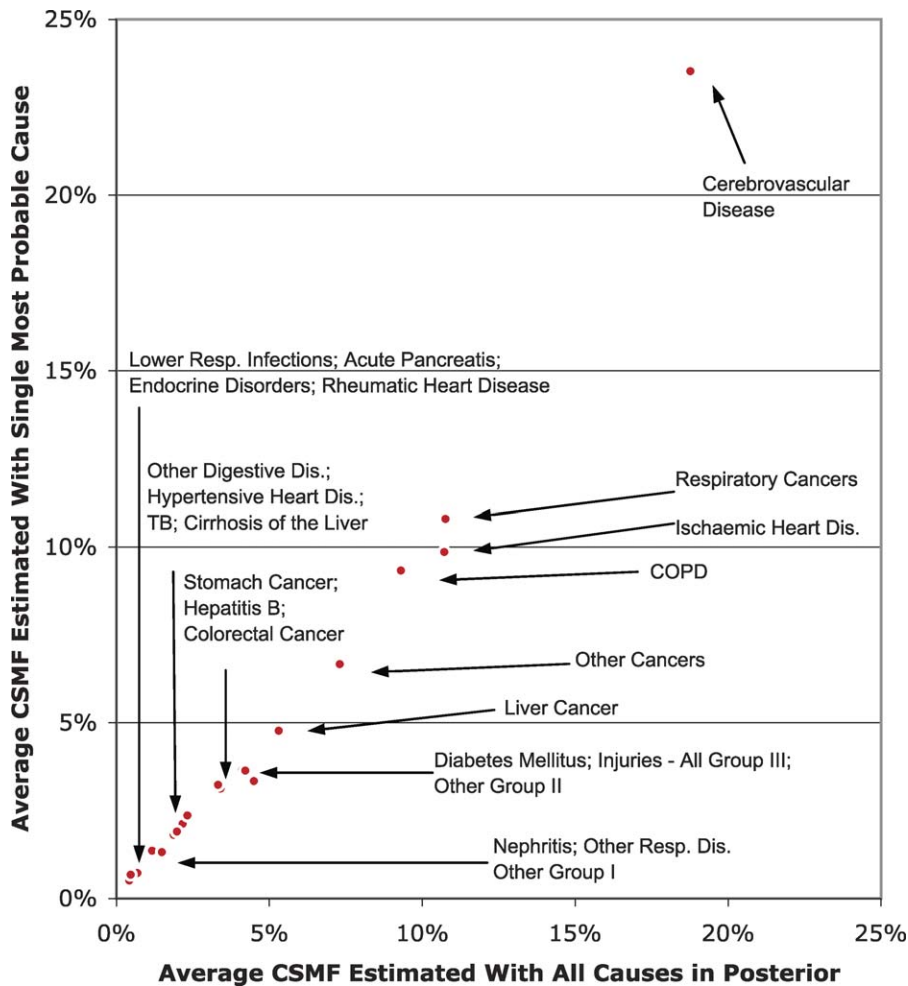
Figure 5. Comparison of the Average CSMF Estimated Using the Single Most Probable Posterior Cause and the Average CSMF Estimated Using the Posterior Distribution of All 23 Causes

The means were computed over the 100 draws of the China data.

doi:10.1371/journal.pmed.0040327.g005

rapidly improve knowledge about countries' epidemiological priorities that would otherwise be impossible from the slow progress of vital registration in developing countries.

Acknowledgments

References

1. Lopez AD, Mathers CD, Ezzati M, Jamison DT, Murray CJL (2006) Global and regional burden of disease and risk factors, 2001: systematic analysis of population health data. Lancet 367: 1747–1757.
2. Khoury SA, Massad D, Fardous T (1999) Mortality and causes of death in Jordan 1995/96: Assessment by verbal autopsy. Bull World Health Organ 77: 641–650.
3. Chen LC, Gesche MC, Ahmed S, Chowdhury AI, Mosley WH (1974) Maternal mortality in rural Bangladesh. Stud Fam Plann 5: 334–341.
4. INDEPTH Network (2003) INDEPTH Standardized Verbal Autopsy questionnaire (Revised August 2003). Available: http://www.indepth-network.org/Resource%20Kit/INDEPTH%20DSS%20Resource%20Kit/INDEPTHVerbalAutopsyInstruments.html. Accessed: 15 October 2007.
5. Jha P, Gajalakshmi V, Gupta PC, Kumar R, Mony P, et al. (2006) Prospective study of one million deaths in India: rationale, design, and validation results. PLoS Med 3: e18. doi:10.1371/journal.pmed.0030018
6. Rao C, Lopez AD, Yang G, Begg S, Ma J (2005) Evaluating national cause-of-death statistics: principles and application to the case of China. Bull World Health Organ 83: 618–625.
7. Black RE, Morris SS, Bryce J (2003) Where and why are 10 million children dying every year? Lancet 361: 2226–2234.
8. Soleman N, Chandramohan D, Shibuya K (2006) Verbal autopsy: current practices and challenges. Bull World Health Organ 84: 239–245.
9. Chandramohan D, Rodrigues LC, Maude GH, Hayes RJ (1998) The validity of verbal autopsies for assessing the causes of institutional maternal death. Stud Fam Plann 29: 414–423.
10. Setel PW, Whiting DR, Hemed Y, Chandramohan D, Wolfson LJ, et al. (2006) Validity of verbal autopsy procedures for determining cause of death in Tanzania. Trop Med Int Health 11: 608–703.
11. Yang G, Rao C, Ma J, Wang L, Wan X, et al. (2006) Validation of verbal autopsy procedures for adult deaths in China. Int J Epidemiol 35: 741.
12. Chandramohan D, Setel P, Quigley M (2001) Effect of misclassification of causes of death in verbal autopsy: can it be adjusted? Int J Epidemiol 30: 509–514.
13. Kalter HD, Gray RH, Black RE, Gultiano SA (1990) Validation of postmortem interviews to ascertain selected causes of death in children. Int J Epidemiol 19: 380–386.
14. Anker M (1997) The effect of misclassification error on reported cause-specific mortality fractions from verbal autopsy. Int J Epidemiol 26: 1090–1096.
15. Maude GH, Ross DA (1997) The effect of different sensitivity, specificity and cause-specific mortality fractions on the estimation of differences in cause-specific mortality rates in children from studies using verbal autopsies. Int J Epidemiol 26: 1097–1106.
16. Murray CJL, Kulkarni SC, Ezzati M (2006) Understanding the coronary

heart disease versus total cardiovascular mortality paradox. A method to enhance the comparability of cardiovascular death statistics in the United States. Circulation 2071–2081.

17. Baiden F, Bawah A, Biai S, Binka F, Boerma T, et al. (2007) Setting international standards for verbal autopsy. Bull World Health Organ 85: 569–648

18. Freeman JV, Christian P, Khatry SK, Adhikari RK, LeClerq SC, et al. (2005) Evaluation of neonatal verbal autopsy using physician review versus algorithm-based cause-of-death assignment in rural Nepal. Paediatr Perinat Epidemiol 19: 323–331.

19. Quigley MA, Chandramohan D, Rodrigues LC (1999) Diagnostic accuracy of physician review, expert algorithms and data-derived algorithms in adult verbal autopsies. Int J Epidemiol 28: 1081–1087.

20. Boulle A, Chandramohan D, Weller P (2001) A case study of using artificial neural networks for classifying cause of death from verbal autopsy. Int J Epidemiol 30: 515–520.

21. Byass P, Lan Huong D, Van Minh H (2003) A probabilistic approach to interpreting verbal autopsies: methodology and preliminary validation in Vietnam. Scand J Public Health 31: 32–37.

22. Quigley MA, Chandramohan D, Setel P, Binka F, Rodrigues LC (2000) Validity of data-derived algorithms for ascertaining causes of adult death in two African sites using verbal autopsy. Trop Med Int Health 5: 33–39.

23. Reeves BC, Quigley M (1997) A review of data-derived methods for assigning causes of death from verbal autopsy data. Int J Epidemiol 26: 1080–1089.

24. Byass P, Fottrell E, Huong DL, Berhane Y, Corrah T, et al. (2006) Refining a probabilistic model for interpreting verbal autopsy data. Sacnd J Public Health 34: 26–31.

25. King G, Lu Y (2007) Verbal autopsy methods with multiple causes of death. Available: http://gking.harvard.edu/files/abs/vamc-abs.shtml. Accessed 17 May 2007.

26. World Health Organization (1993) International Statistical Classification of Diseases and Related Health Problems, 10th Revision, 2nd Ed. Geneva: WHO.

27. Rao C, Yang G, Hu J, Ma J, Xia W, et al. (2007) Validation of cause-of-death statistics in urban China. Int J Epidemiol 36: 642–651.

28. Kahn K, Tollman SM, Garenne M, Gear JSS (2000) Validation and application of verbal autopsies in a rural area of South Africa. Trop Med Int Health 5: 824–831.

29. Setel PW, Rao C, Hemed Y, Whiting DR, Yang G, Chandramohan D, Alberti K, Lopez AD (2006) Core verbal autopsy procedures with comparative validation results from two countries. PLoS Med 3: e268. doi:10.1371/journal.pmed.0030268

30. Varmus H, Klausner R, Zerhouni E, Acharya T, Daar AS, et al. (2003) Public health. Grand challenges in global health. Science 302: 398–399.

31. Lopman BA, Barnabas RV, Boerma JT, Chawira G, Gaitskell K, et al. (2006) Creating and validating an algorithm to measure AIDS mortality in the adult population using verbal autopsy. PLoS Med 3: e312. doi:10.1371/journal.pmed.0030312

32. Urassa M, Boerma JT, Isingo R, Ngalula J, NgÆweshemi J, et al. (2001) The impact of HIV/AIDS on mortality and household mobility in rural Tanzania. AIDS 15: 2017–2023.

33. Murray CJ, Lopez AD (1997) Global mortality, disability, and the contribution of risk factors: Global Burden of Disease Study. Lancet 349: 1436–1442.

Editors' Summary

Background. All countries need to know the leading causes of death among their people. Only with accurate cause-of-death data can their public-health officials and medical professionals develop relevant health policies and programs and monitor how they affect the nation's health. In developed countries, vital registration systems record specific causes of death that have been certified by doctors for most deaths. But, in developing countries, vital registration systems are rarely anywhere near complete, a situation that is unlikely to change in the near future. An approach that is being used increasingly to get information on the patterns of death in poor countries is "verbal autopsy" (VA). Trained personnel interview household members about the symptoms the deceased had before his/her death, and the circumstances surrounding the death, using a standard form. These forms are then reviewed by a doctor, who assigns a cause of death from a list of codes called the International Classification of Diseases. This process is called physician-coded verbal autopsy (PCVA).

Why Was This Study Done? PCVA is a costly, time-consuming way of analyzing VA data and may not be comparable across regions, because it relies on the views of local doctors about the likely causes of death. In addition, although several studies have suggested that PCVA is reasonably accurate, such studies have usually included information collected from household members about medical records or contacts with health services. In regions where there is little contact with health services, PCVA may be much more inaccurate. Ideally what is needed is a method for assigning causes of death from VA data that does not involve physician review. In this study, the researchers have developed a statistical method—the symptom pattern (SP) method—for analyzing VA data and asked whether it can overcome the limitations of PCVA.

What Did the Researchers Do and Find? The SP method uses VA data collected about a group of patients for whom the true cause of death is known to calculate the probability for each cause of death that a household member will answer yes when asked about various symptoms. These so-called "symptom properties" can be used to calculate population cause-specific mortality fractions (CSMFs—the proportion of the population that dies from each disease) from VA data and, using a type of statistical analysis called Bayesian statistics, can be used to assign causes of deaths to individuals. When used with data from a VA study done in China, the SP method estimated population CSMFs with an average relative error of 16% (this measure indicates how much the estimated and true CSMFs deviate), whereas PCVA estimated them with an average relative error of 27%. At the individual level, the SP method assigned the correct cause of death in 83% of cases; PCVA was right only 69% of the time. Removing the medical record recall section of the VA data had little effect on the accuracy with which the two methods estimated population CSMFs. However, whereas the SP method still assigned the correct cause of death in 78% of individual cases, the PVCA did so in only 38% of cases

What Do These Findings Mean? These findings suggest that the SP method for analyzing VA data can outperform PCVA at both the population and the individual level. In particular, the SP method may be much better than PCVA at assigning the cause of death for individuals who have had little contact with health services before dying, a common situation in the poorest regions of world. The SP method needs to be validated using data from other parts of the world and also needs to be tested in multi-country validation studies to build up information about how culture and language affect the likelihood of specific symptoms being reported in VAs for each cause of death. Provided the SP method works as well in other countries as it apparently does in China, its adoption, together with improvements in how VA data are collected, has the potential to improve the accuracy of cause-of-death data in developing countries.

Additional Information. Please access these Web sites via the online version of this summary at http://dx.doi.org/10.1371/journal.pmed.0040327.

- An accompanying paper by Murray and colleagues describes an alternative approach to collecting accurate cause-of-death data in developing countries
- World Health Organization provides information on health statistics and health information systems, on the International Classification of Diseases, on the Health Metrics Network, a global collaboration focused on improving sources of vital statistics and cause-of-death data, and on verbal autopsy standards
- Grand Challenges in Global Health provides information on research into better ways for developing countries to measure their health status