ORIGINAL RESEARCH

Multiple causes of death models for human immunodeficiency virus/acquired immune deficiency syndrome and related mortality in South Africa in 2006 and 2007

Alphonse KA Amey[1]
Ntonghanwah Forcheh[2]
Keamogetse Setlhare[2]

[1]Departments of Statistics, University of Venda; [2]South Africa University of Botswana, Gaborone, Botswana

Background: Predictive models for mortality due to human immunodeficiency virus (HIV) disease as a result of opportunistic infections, such as tuberculosis and pneumonia, have been developed.

Methods: The data are taken from the Statistics South Africa multiple causes of death data for 2006 and 2007, which is available for public use. The dataset was compiled from death notifications, and contains up to five causes of death as well as some demographic characteristics of the deceased. The logistic regression modeling framework was used to model the presence or absence of HIV disease, given the predictive variables.

Results: The higher the number of causes listed, the higher the likelihood that HIV would be a cause, with the percentage of notifications of HIV listed increasing from under 2% when only one cause is listed to almost 15% when 4–5 causes are listed. When the logit model was fitted to the multiple cause of death model, it was found that individual demographics were good predictors of the likelihood that the death notification would have HIV as one of the causes of death. Although there are highly significant differences in the likelihood that people of different demographics would die from HIV, the predictive power of these demographic factors on their own is very low, especially when there is only a single cause of death mentioned. With the full multiple cause of death model, two-way interactions between tuberculosis, pneumonia, and other opportunistic infections were highly significant, and their inclusion lead to significant improvements in the predictive power of the model.

Keywords: multiple causes of death, underlying causes of death, human immunodeficiency virus, logit model, odd ratios

Introduction

It has long been acknowledged that death rarely arises from a single cause. Weiner et al[1] examined death notifications that recorded multiple causes of death and found an increase in the number of deaths associated with hypertension compared with the number obtained if only the underlying cause of death was recorded. Israel et al[2] noted that some causes of death, such as ill-defined conditions, are usually recorded as underlying causes only when they appear alone in the death notification. Furthermore, what is recorded as the underlying cause may be influenced by how the attending doctor views the diseases present at the time of death. This observation is particularly true for non-autopsied deaths. Santo et al[3] argue that although there has been increased registration of tuberculosis, its importance to populationwide mortality is not completely

Correspondence: Alphonse KA Amey
Department of Statistics, University of Venda, Private Bag X5050, Thohoyandou 0950, South Africa
Tel +27 15 062 8437
Fax +27 86 624 3234
Email alphonse.amey@univen.ac.za

1

reflected by statistics accompanying underlying cause of death. They noted that tuberculosis occurs as an associated cause in a large number of death notifications in which the demise is attributed to a different underlying cause.

Notwithstanding these concerns, only a few countries, such as the US, Canada, and the UK, undertook multiple cause of death (MCOD) coding prior to introduction of the Tenth Revision of the International Classification of Diseases (ICD-10). In addition to recording the underlying (main) cause of death, national databases that use the ICD-10 classification include other immediate causes which may or may not be the same as the underlying cause, as well as other conditions present or associated with the underlying cause.[1,4,5] Although it has been noted that not all conditions present at the time of death are necessarily recorded on death notifications,[1] multiple causes of death data provide researchers with a more comprehensive history of long-term illness which do not often appear as the underlying cause of death.[5]

Researchers differ considerably on the impact of human immunodeficiency virus (HIV)/acquired immunodeficiency syndrome (AIDS) on mortality owing to many challenges in reporting and diagnosing HIV prevalence and mortality due to AIDS.[6] Estimates of the number of people with HIV or dying from AIDS also differ considerably, especially in developing countries.[7] Although much research has been conducted on HIV/AIDS, few of the studies have actually utilized nationally representative datasets for estimating HIV/AIDs mortality. Furthermore, even when national datasets have been utilized, these have been aggregated data in which only underlying cause of death is indicated.[7] Multiple causes of death data provide a wealth of information to conduct more indepth research on the contribution of HIV/AIDS to mortality.

AIDS was first included as one of the causes of death in Revision 9 of the International Classification of Diseases (ICD-9). The underlying cause of death was taken to be AIDS if HIV was present along with more than one opportunistic disease/condition, ie, the so-called AIDS defining condition. In ICD-10, AIDS has been replaced by several groups of causes, called HIV diseases. The link between tuberculosis, AIDS, and other related causes of death has been of interest to both health practitioners and researchers, although few studies appear to have been done at a national level to corroborate this association.

The Southern African region is considered to be among the areas of the world most afflicted by HIV/AIDS, and not surprisingly, the highest differences about the perceived impact of HIV/AIDS on morbidity and mortality prevail in this region. Using the 2006 and 2007 multiple causes of death data for South Africa, this paper develops models for predicting mortality due to HIV diseases, given sociodemographic factors, as well as the presence/absence of tuberculosis, pneumonia, and other opportunistic infections.

The methodology for analyzing MCOD data is still in its infancy. Bah[8] compared the profile of leading causes of death in South Africa from 1997 to 2001 using single cause of death data to profile multiple causes of death data. Bah observed that analysis of multiple cause of death data de-emphasized the importance of ill-defined conditions, and emphasized the greater prominence of tuberculosis, pneumonia, and influenza, and called for further research to study the association between the different causes of death. Redelings et al[9] highlighted some of the challenges in trying to model disease causality using data from death notifications. These challenges arise from physicians' beliefs regarding which of the causes were underlying and which were associated; omission from notifications of conditions that may be present but considered unrelated to death and researchers' lack of knowledge on what percentage of notification should contain any pair of causes purely by chance. Despite the challenges, Redelings et al[9] concluded their study by insisting that multiple causes of death data provide an easily available source of population information on causes of death and associated causes of death. Furthermore, they propose that multiple causes of death data should be analyzed by looking at the frequencies with which pairs of conditions (such as tuberculosis and pneumonia) co-occur. However, they caution that it is not clear as to whether the observed percentage of death with comorbidity is different from what we expect. For example, for the data used in their study, 40.8% of notifications that reported hypertension as a cause of death also reported heart disease as a cause, but it is not clear if this is what we expect or not.

Wall et al[10] studied the relationship between MCOD as a function of factors listed on Minnesota death notifications from 1990 to 1998. They computed and compared the frequency with which a given cause was listed as an MCOD and frequency with which it was listed as an underlying cause for different demographic groups. They used logistic regression to determine the odds of having a particular condition, such as ischemic heart disease, as underlying rather than a contributing cause given the demographic background.

Santo et al[3] limited their study to computing death rates due to tuberculosis as an underlying cause and as a mentioned cause. They compared these rates among males and females and among different age groups. They also determined the

proportion of notifications in which deaths were due to tuberculosis that contain different associated causes, such as pneumonia and respiratory failure. AIDS was the main underlying cause in 65% of the deaths in which tuberculosis was mentioned as an associated cause. In the group aged 20–49 years, this percentage was 86%. They suggested that coinfection by tuberculosis and AIDS is an important factor in premature mortality.

This paper focuses on all deaths for which HIV/AIDS or other opportunistic infections associated with HIV/AIDS were mentioned as a cause. Apart from HIV/AIDS, the leading causes of mortality in Southern Africa, such as tuberculosis, pneumonia, ill-defined conditions, and respiratory infections have been around for a long time, and hence have been studied using the single cause data framework.

Materials and methods

The data for this study are taken from the 2006 and 2007 South African multiple causes of death data. Statistics South Africa (Stats-SA) has generously provided these data to researchers to use free of charge for nonprofit purposes, with the simple request that Stats-SA is acknowledged as the original source of the data. Full details of the compilation and coding process used to obtain the South African MCOD data are presented in the meta file for each year's data.

The South African Births and Deaths Registration Act of 1992 requires that all deaths and stillbirths occurring in the country be registered using the notification/register of death/stillbirth form (B1-1663). Stats-SA collects all these notifications for each year and uses a comprehensive set of procedures to process the forms and extract the causes of death from the death notifications. The 2006 and 2007 South African causes of death data were coded using ICD-10 codes, while the underlying cause of death was derived automatically using a United States National Center for Health Statistics software program called Automated Classification of Medical Entities (ACME 2000.05) that applies World Health Organization rules to the selection of underlying causes of death.[11]

Deaths that were not registered at the Department of Home Affairs and deaths registered late or registered but had not reached Stats-SA during the processing phase for a given year are not included.[12] Anderson and Philips[13] estimated that about 67% of deaths were registered in South Africa in 1996, and that this percentage has ranged from 78% to 84% since 1997. They further estimate that death registration is most complete among the population aged 65 years and over, and least complete among children, with only 40%–57% of

deaths occurring among children aged 0–14 years registered annually since 1997.

Data coding and transformation

The South African MCOD data provides up to five possible causes of death to be specified, and these are recorded as cause A, B, C, D, and "other cause". Cause A represents the immediate condition leading to death, while if present, causes B, C, and D represented sequentially, the second, third, and fourth conditions leading to death (Stats-SA data for 2006 and 2007). "Other significant conditions contributing to death but not resulting in the underlying cause" are recorded as "other cause" while "the disease or injury which initiated the train of morbid events leading directly to death, or the circumstances of the accident or violence which produced the fatal injury" constitutes the underlying cause of death. This underlying cause is selected from among causes A–D. The dataset also includes the broad and main groups for the underlying cause.

The broad groups of causes of death that are of interest in this study and the specific ICDS-10 codes are as follows: HIV diseases, ICDS-10 codes B20–B24 and R75; tuberculosis, A15–A19, B90, U51, U52; pneumonia, J12–J18, P23, B59; other opportunistic diseases, excluding tuberculosis and pneumonia, consisting of diarrhea (A09), candidiasis (B37), cytomegalovirus (B25), non-Hodgkin's lymphoma (C82, C85), cervical cancer (C53, D06), coccidiomycosis (B38), herpes simplex (B00, B02, A60), meningitis (A39, A87, G00, G03), cryptococcosis (B45), histoplasmosis (B39), toxic encephalopathy (G92), Kaposi's sarcoma (C46), toxoplasmosis (B58), cryptosporidiosis (A07), and Salmonella septicemia (A02).

The dataset includes the following background of the deceased when known: gender, marital status, province of birth, province of death, institution of death, province of residence, education, usual occupation, type of business/industry, smoking status, and month and year of death. From these, death institution, gender, age group, marital status, and education were included in the models. Smoking status, occupation, and industry were not included because of a high percentage of missing values. It was desirable to include racial classification since certain causes of morbidity and mortality are known to differ by racial groups,[14] but this information was not captured in the electronic data.

The background factors from the death notification used in the analysis and their coding are:

X1)   age group (0, under 5 years; 1, 5–14 years; 2, 15–24 years; 3, 25–49 years; 4, 50–64 years; 5, 65+ years)
X2)   gender (0 male, 1 female)

X3) level of education (0, none; 1, primary; 2, secondary; 3, post-secondary; 4, unknown)

X4) Year of death (1, 2006; 0, 2007)

X5) Death institution (0, hospital; 1, emergency room, outpatient or death on arrival; 2, home; 3, other/unspecified)

X6) Marital status (0, single; 1, married; 2, widowed; 3, divorced/separated; 4, not known).

In addition, some of the models include the total number of causes listed (TC).

With MCOD data, the contribution of a particular disease to death could be expressed using a number of alternative categorical variables. The simplest case is to specify whether the disease was an underlying cause or not. An alternative is to indicate whether or not the disease was mentioned (as underlying, immediate, other condition leading to death, or other significant condition contributing to death but not resulting from the underlying cause). Yet a third option is to create a categorical variable with the following categories: 0, not mentioned; 1, mentioned but not underlying; and 2, mentioned as underlying only.

Modeling

To obtain models that allow for comparison of MCOD with a single cause of death (SCOD) analysis, the following coding of each listed cause, which retains the importance of underlying cause used in single cause of death analyses, was of interest: 0, cause not mentioned on notification; 1, cause mentioned but not underlying; and cause mentioned as underlying. This coding leads to an ordinal response variable in terms of the importance of the disease as a cause of death.

Exploratory analysis of HIV-related deaths in the dataset used in this paper revealed that very few notifications had HIV disease mentioned as a cause but not as underlying. Therefore, it was not possible to treat the presence of HIV disease on a death notification as an ordinal variable. The only option was to model the presence ($Y = 1$) or absence ($Y = 0$) of HIV on the death notifications. The logit model for the presence of HIV (y = 1) given a vector of predictive factors, x, is

$$\Pr(y_i = 1 / \mathrm{x}_i) = P_i = \frac{\exp(\mathrm{x}'_i \beta)}{1 + \exp(\mathrm{x}'_i \beta)} = \frac{\eta_i}{\eta_{i0} + \eta_i},$$

Indicator variables denoting the presence or absence of tuberculosis, pneumonia, and other opportunistic infections respectively (z), together with background factors (x) are used as explanatory variables in the model. We refer to the model involving only the background factors, with (x) as the background factors model. The background factors model

can be fitted to the SCOD as well as the MCOD data. When at least one other HIV-associated cause of death (tuberculosis, pneumonia, or other opportunistic infection) is mentioned, then the full multiple cause of death model is given by:

$$\Pr[y = 1|(x, z)] = \exp(\alpha' x + \beta' z)/[1 + \exp(\alpha' x + \beta' z)].$$

With this model, the odds of HIV-disease being mentioned on a death notification of a deceased person with background characteristics vector **x**, and associated causes vector **z**, can be examined for various combinations of background factors and associated causes.

Results

As shown in Table 1, the dataset consisted of 1,236,690 cases, with 621,378 for 2006 and 615,312 for 2007. Of these, 15,538 cases had HIV listed as a cause of death in 2006 and 14,380 in 2007, giving a total of 29,918 (2.4%) cases for the two years. There are 461,568 (37.3% of all deaths recorded in 2006 and 2007) cases for which HIV/AIDS or other opportunistic infections associated with HIV/AIDS (Table 1) were mentioned as a cause. A total of 748,328 (60.5%) notifications had just a single cause of death mentioned, and a further 26.6% had just two causes listed.

Figure 1 shows the percentage of death notifications that contained HIV disease as a cause by the number of causes listed on the notification. Very few of single cause deaths (1.9%) are listed as due to HIV disease. The proportion of notifications with HIV listed as one of the causes increases steadily with the number of causes listed, reaching a peak of 15.3% for combined males and females. These percentages are consistently higher for female death notifications than for males.

Figure 2 shows the percentages of male and female death notifications that had HIV, tuberculosis, pneumonia, and other opportunistic infections listed among the causes by age group. The majority of death notifications with HIV listed as a cause increases among females up to 49 years of age, and then declines (Figure 2A). A similar pattern is observed among males, except for the age group 15–24 years who have a lower percentage. From the age group 15–64 years, proportionately more female death notifications include HIV. The percentage of death notifications for which tuberculosis is listed as a cause is almost the same among males and females aged 0–24 years (Figure 2B). Thereafter, the percentage for males is higher than for females.

The percentage of death notifications with pneumonia listed is highest among the younger and older age groups

Table 1 Deaths for which HIV is listed as a cause (complete data)

	Year	Number of causes listed on death notification					Total
		1	2	3	4	5	
All deaths	Total	748,328	329,081	118,395	32,847	8039	1,236,690
	2006	375,149	166,027	59,480	16,630	4092	621,378
	2007	373,179	163,054	58,915	16,217	3947	615,312
All deaths without HIV listed	Total	743,957	314,090	110,860	30,453	7412	1,206,772
	2006	372,635	158,496	55,574	15,400	3735	605,840
	2007	371,322	155,594	55,286	15,053	3677	600,932
All deaths with HIV listed	Total	4371	14,991	7535	2394	627	29,918
	2006	2514	7531	3906	1230	357	15,538
	2007	1857	7460	3629	1164	270	14,380
All deaths without HIV but at least one of TB, pneumonia, or other OI listed	Total	228,854	134,417	50,534	14,383	3462	431,650
	2006	116,259	68,844	25,466	7262	1712	219,543
	2007	112,595	65,573	25,068	7121	1750	212,107
All deaths with at least one of HIV, TB, pneumonia, or other OI listed	Total	233,225	149,408	58,069	16,777	4089	461,568
	2006	118,773	76,375	29,372	8492	2069	235,081
	2007	114,452	73,033	28,697	8285	2020	226,487

Abbreviations: HIV, human immunodeficiency virus; OI, opportunistic infection; TB, tuberculosis.

(Figure 2C) and lowest among those aged 15–49 years. Other opportunistic infections are frequently listed among the group under five years of age (Figure 2D) and decline steadily for both male and female death notifications. For the age group 25 years and over, relatively more female death notifications than male notifications include other opportunistic infections.

Background factor models for all deaths

In this first model, the odds that HIV is listed as a cause of death, whether, immediate, underlying, secondary, or other is modeled using demographic variables and TC as predictors.

The test that all coefficients of the predictors are simultaneously equal to zero has a chi-square value of 23419.723, with 24 degrees of freedom and an observed P value < 0.001. All the demographic variables were significant

in predicting the odds that HIV disease is listed as a cause of death (Appendix A). In particular, the odds depend on the gender-age interaction and the TC listed. As shown in Figure 3, females aged 25–49 years have a significantly higher likelihood of having HIV listed than their male counterparts. From 50 years onwards, the odds for females remain slightly higher than for males, while there are no differences among people aged 24 years or younger.

The model results further show that each additional cause of death listed doubles the odds that HIV will be one of the causes of death. Hence, those for whom HIV disease is listed as a cause of death are more likely to have multiple causes listed than from HIV alone, even when the other causes are not HIV-related. However, the predictive validity of this model is low. While the model can be used to reclassify all deaths correctly for which HIV is not listed as such, only 0.3% of deaths with HIV listed as a cause of death are correctly classified using this model.
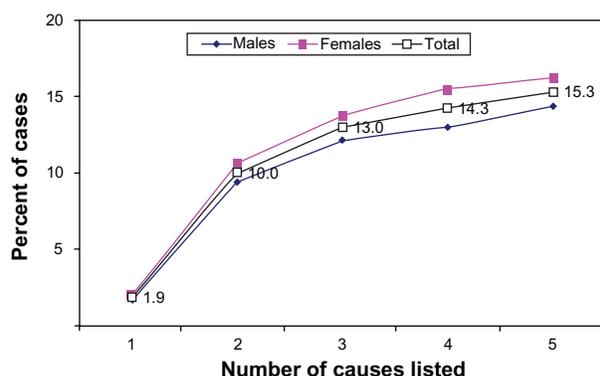
Models for a single cause of death listed on death notification

The data consisted of 233,225 deaths in which only one of the four HIV-related causes (HIV, tuberculosis, pneumonia, or other opportunistic infection) was listed. As shown in Figure 1, only 1.9% (4371) of these deaths was attributed to HIV disease. For these data, only background factors could be included as predictors of the likelihood that the death would be due to HIV disease. For this model, the overall model chi-square is 2010.86 with 23 degrees of freedom and an observed P value < 0.001.

Figure 1 Percentage of human immunodeficiency virus deaths by number of causes listed and gender of deceased.

A HIV is listed as a cause of death

B TB is listed as a cause of death

C Pneumonia is listed as a cause of death

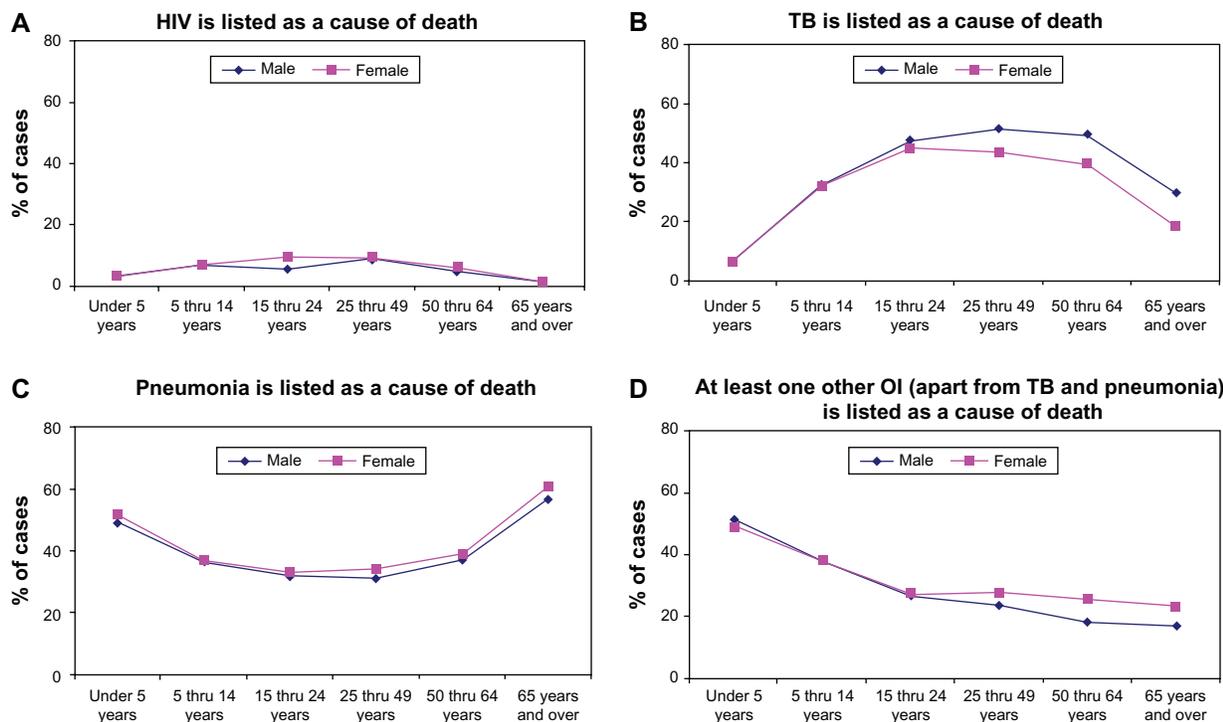**D** At least one other OI (apart from TB and pneumonia) is listed as a cause of death

Figure 2 Deaths due to human immunodeficiency virus (HIV), tuberculosis (TB), pneumonia, and other opportunistic infections (OIs).
Abbreviation: thru, through.

The results indicate that the odds of HIV being listed as the single cause of death depend on death institution, age group, marital status and education, but not on gender (Appendix B). Deaths for which HIV is listed as the single cause are significantly less likely to occur in hospitals/nursing homes than elsewhere. The odds of deaths occurring in emergency rooms/on arrivals/outpatients are twice the odds of death occurring in hospitals/nursing homes, and 85% more likely for deaths occurring at home.

However, the predictive power of the model was very low. When the model is used to reclassify deaths as due to HIV, it performs very poorly. None of the 1.9% of deaths for which HIV is a listed cause is correctly reclassified as such,
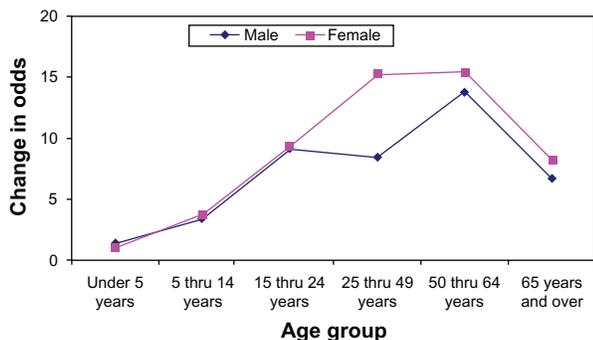
Figure 3 Odds of having human immunodeficiency virus listed as a cause of death: gender–age interaction.
Abbreviation: thru, through.

while all the 98.1% of cases in which HIV is not a listed cause were correctly reclassified as such. Since the different factors are significant, the poor performance of the model is due mainly to the fact that HIV deaths are rare among notifications with only one cause of death listed.

Models for multiple causes of death listed on death notification

Three classes of models were fitted for notifications, corresponding with two, three, and four or more diseases listed on the death notifications. This is because if k-causes are listed, then only k-1 causes can be included in any model for predicting the likelihood that HIV will be listed as a cause of death. Therefore, to avoid the problem of introducing structural zeroes, when a single cause is listed, only the background factors model can be fitted: when two causes are listed, then each of the HIV-related causes (HIV disease, tuberculosis, pneumonia, and other opportunistic infection) must be included alone in any model; when three causes are listed, any pair of the three causes can be included in the model; and while four or more causes are listed, then all three related causes can be included in a single model.

Models for notifications with two causes of death listed

Of the 149,408 notifications for which exactly two HIV-related causes of death were listed, 14,991 (10.0%) had

HIV listed as one of the two causes of death. For these cases, it was not possible to fit a model with more than one other cause as a predictor of HIV. Three separate models were therefore considered with background factors, plus tuberculosis as an additional predictor, pneumonia as a predictor, and other opportunistic infections as predictors.

For the model in which tuberculosis was included along with background factors as predictors, the results show that the odds of HIV being listed as a cause of death depends on death institution, age, marital status, education, year of death, and whether or not tuberculosis is listed as one of the causes (Appendix C). The significance of gender in the model is only through its interaction with age. When tuberculosis is listed as a cause of death, the odds of HIV also being listed reduces significantly by about 23%. That is, when only two causes are listed on a notification, then notifications with HIV and tuberculosis jointly listed were less likely than those with only one of them listed.

When tuberculosis was replaced by pneumonia in the model, the significant factors were found to be death institution, age, marital status, education, and the gender-age interaction (Appendix D). When pneumonia is listed as a cause of death, the odds of HIV also being listed reduces significantly by about 58%.

When pneumonia was replaced by other opportunistic infections, the odds of HIV being listed on a notification with only two listed causes as a cause of death depends on death institution, age, marital status, education, and the gender-age interaction (Appendix E). Furthermore, the odds of HIV being listed along with some other opportunistic infection reduces significantly by about 55%.

These models indicate that when only two causes are listed, there is a significant reduction in the odds of HIV being listed along with any associated causes, ie, tuberculosis, pneumonia, or some other opportunistic infections.

Models for notifications with three causes of death listed

When three causes are listed, it is possible to fit models with two of the three HIV-associated causes. Hence three different models were fitted. The first model included tuberculosis and pneumonia as predictors, the second included tuberculosis and other opportunistic infection, while the third model included pneumonia and other opportunistic infection as predictors of the likelihood that HIV would be listed as a cause of death. There are 58,069 cases in the dataset for which exactly three causes of death are listed on the death notification. Of these, 7535 cases (13.0%) had HIV listed as a cause of death.

In all three models, sociodemographic variables are strong predictors of the odds of HIV being listed as a cause (Appendices F–H). The significance of gender in the models is through its interaction with age. Significant interactions also exist between tuberculosis, pneumonia, and opportunistic infection, as shown in Figure 4A–D.

Figure 4A–C depict the interactions between tuberculosis and pneumonia, tuberculosis and other opportunistic infection, and pneumonia and other opportunistic infection, respectively. In Figure 4A, the highest odds of HIV being listed are when neither tuberculosis nor pneumonia is listed, followed by the odds when tuberculosis is listed but not pneumonia. When tuberculosis is not listed, the odds of HIV being listed when pneumonia is present is lower (by about 63%) than when pneumonia is not present. When pneumonia is listed, the odds do not depend on whether or not tuberculosis is listed. As shown, the interaction is even stronger between tuberculosis and other opportunistic infections. When tuberculosis is listed, the odds of HIV being listed remain almost constant, whether or not some other opportunistic infection (other than pneumonia) is listed (Figure 4B). However, when tuberculosis is not listed, the odds of HIV being listed are lower (by about 46%) when some other opportunistic infection (other than pneumonia) is listed than when no other opportunistic infection is listed. In Figure 4C, when pneumonia is listed, the odds of HIV also being listed is approximately unchanged, whether some other opportunistic infection (other than tuberculosis) is listed or not. However, when pneumonia is not listed, the odds of HIV being listed are significantly lower (by about 68%) when some other opportunistic infection (other than tuberculosis) is listed than when no other opportunistic infection is listed.

Models for notifications with four or more causes of death listed

Of 20,866 cases with more than three causes of death listed, 3021 (14.5%) had HIV as one of the causes of death. The full model with all three related causes of death was fitted including relevant interaction terms (Table 2). Among the background factors, education of the deceased was not a significant factor in predicting HIV as a cause of death, while the interaction between age and gender was significant. The two-way interactions among tuberculosis, pneumonia, and opportunistic infection were also significant. For cases with tuberculosis and pneumonia not listed, the odds of HIV being listed as a cause of death are significantly lower when other opportunistic infections are listed. When tuberculosis
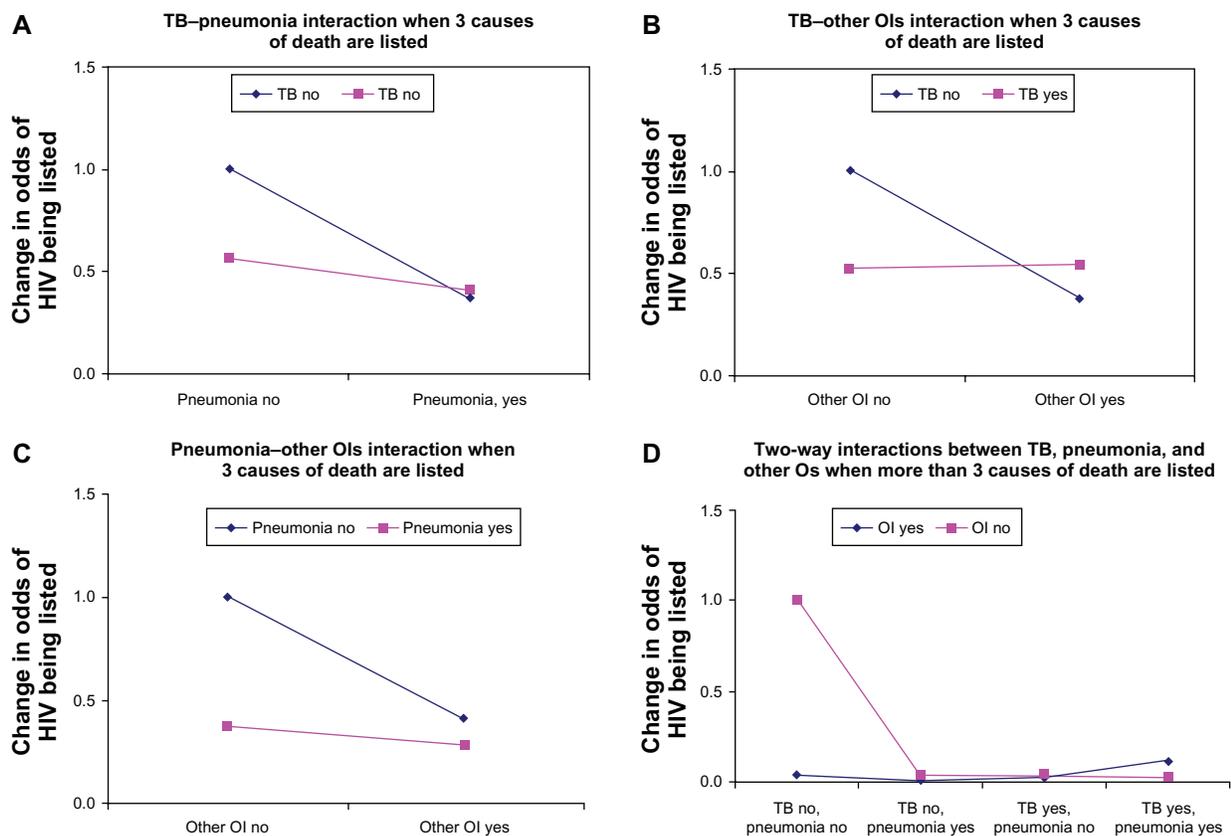
Figure 4 Interaction among tuberculosis (TB), pneumonia, and other opportunistic infections (OIs).
Abbreviation: HIV, human immunodeficiency virus.

and pneumonia are both listed, the odds of HIV being listed is higher for cases with opportunistic infection listed than cases with opportunistic infection not listed. The full nature of the two-way interactions is shown in Figure 4D.

Discussion

This study found that individual demographics were good predictors of the likelihood that the death notification would have HIV as one of the causes of death. However, the predictive power of the background factors model for single cause of death data is very low. This is mainly because deaths in which HIV is the only cause are rare even when one considers only notifications in which only one of HIV, tuberculosis, pneumonia, and other opportunistic infection are listed as the single cause of death. Hence no background factors can adequately predict whether any particular individual is more/less likely to die of HIV disease based only on their gender, age, marital status, education, and death institution. This makes multiple cause of death analysis imperative.

The different models for multiple causes of death reveal interaction between age and gender of the deceased as an important determinant of the likelihood that death would be

caused by HIV disease. Females aged 25–49 years are much more likely to have HIV listed as a cause of death than males in the same age group. For the younger age groups, there is no difference between males and females, while for older age groups, females remain slightly more likely than men to have HIV listed as one of the causes. The case for older age groups is rather surprising given that intergenerational sex, which is believed to be one of the factors responsible for HIV infection among the older population, is more prominent between older men and younger women than between older women and younger men.

The results also show that the more the number of causes listed, the higher the likelihood that HIV would be a cause, with the percentage of notifications with HIV listed increasing from under 2% when only one cause is listed to almost 15% when 4–5 causes are listed. However, despite this increase, the likelihood of HIV being listed as a cause of death actually decreases when tuberculosis and/or pneumonia are also listed.

Various two-way interactions between tuberculosis, pneumonia, and other opportunistic infections reveal that the decrease in the odds between the base model (when none of the other factors is present) and other models (when one or

more of the other factors are present) depends on which two other causes are co-listed.

The main thrust of the results for full multiple causes of death models as shown in Figure 4 and Table 2 reveal that when opportunistic infections other than tuberculosis are listed, the odds of HIV also being listed drops significantly (almost zero). When no other opportunistic infections apart from tuberculosis and pneumonia are listed, the odds of HIV being listed also drops significantly, compared with when only some other opportunistic infection is listed and

not tuberculosis nor pneumonia. It appears that tuberculosis, pneumonia, and other opportunistic infections are being used as proxies for HIV on death notifications!

The performance of the models improves significantly with the addition of related causes of death (tuberculosis, pneumonia, and other opportunistic infections). Hence, although factors such as age, gender, education, marital status, level of education, and death institution are significant predictors of HIV-related deaths, their predictive power is quite low even when considering only deaths due to HIV disease and related causes.

Table 2 Full multiple causes of death model for predicting likelihood of human immunodeficiency virus being listed as a cause of death given background factors and presence or absence of tuberculosis, pneumonia, and other opportunistic infections.

	B	SE	Wald	df	Sig	Exp (B)	95% CI exp (B)	
							Lower	Upper
Death year (2006)	−0.13	0.044	8.1	1	0.004	0.88	0.81	0.96
Death institution			128.8	3	0.000			
0 hospital or nursing home	0					1		
1 emergency room, outpatient, or dead on arrival	−0.26	0.132	3.9	1	0.049	0.77	0.60	1.00
2 home	−1.02	0.091	125.2	1	0.000	0.36	0.30	0.43
3 other/unspecified	0.01	0.071	0.0	1	0.867	1.01	0.88	1.16
Gender (female)	0.30	0.113	7.0	1	0.008	1.35	1.08	1.68
Age, years			170.5	5	0.000			
0, under 5	0					1		
1, 5–14	0.74	0.206	12.8	1	0.000	2.09	1.39	3.13
2, 15–24	0.16	0.254	0.4	1	0.525	1.18	0.72	1.93
3, 25–49	0.64	0.145	19.1	1	0.000	1.89	1.42	2.51
4, 50–64	−0.12	0.167	0.5	1	0.483	0.89	0.64	1.23
5, 65+	−1.40	0.226	38.4	1	0.000	0.25	0.16	0.38
Education			5.6	4	0.230			
0 none	0					1		
1 primary	0.04	0.125	0.1	1	0.779	1.04	0.81	1.32
2 secondary	0.02	0.122	0.0	1	0.902	1.02	0.80	1.29
3 post-secondary	−0.30	0.24	1.6	1	0.207	0.74	0.46	1.18
4 not known	−0.08	0.115	0.5	1	0.472	0.92	0.73	1.15
Age, years (gender)			14.7	5	0.012			
5–14 (female)	−0.34	0.267	1.6	1	0.204	0.71	0.42	1.20
15–24 (female)	0.49	0.259	3.6	1	0.058	1.63	0.98	2.71
25–49 (female)	−0.16	0.126	1.7	1	0.196	0.85	0.66	1.09
50–64 (female)	−0.12	0.177	0.4	1	0.504	0.89	0.63	1.26
65+ (female)	−0.79	0.311	6.5	1	0.011	0.45	0.25	0.83
Marital status			19.9	4	0.001			
1 single	0					1		
2 married	−0.15	0.068	5.1	1	0.024	0.86	0.75	0.98
3 widowed	−0.36	0.166	4.6	1	0.033	0.70	0.51	0.97
4 divorced	−0.03	0.215	0.0	1	0.887	0.97	0.64	1.48
5 not known	0.16	0.064	6.0	1	0.015	1.17	1.03	1.32
Tuberculosis	−3.18	0.103	957.1	1	0.000	0.04	0.03	0.05
Pneumonia	−3.63	0.105	1202.6	1	0.000	0.03	0.02	0.03
Other opportunistic infections	−3.37	0.104	1049.4	1	0.000	0.03	0.03	0.04
Tuberculosis by pneumonia	2.98	0.115	667.4	1	0.000	19.66	15.68	24.64
Tuberculosis by other opportunistic infections	2.90	0.115	636.5	1	0.000	18.18	14.51	22.78
Pneumonia by other opportunistic infections	2.10	0.116	327.7	1	0.000	8.13	6.48	10.20
Constant	1.36	0.128	113.8	1	0.000	3.90		

Abbreviations: CI, confidence interval; df, degrees of freedom; Sig, significance; SE, standard error of the mean.

For the full multiple cause of death model, the only nonsignificant factor here is education. This indicates that HIV does not discriminate by education. The full model is still far from accurately classifying HIV deaths, given these highly significant predictors. While the model correctly predicts 99.9% of cases for which HIV is not listed as a cause, it only correctly predicts 20.7% of cases for which HIV is listed as a cause. This gives an overall reclassification rate of 88.6%. Furthermore, the Cox and Snell multiple correlation coefficient is 13.5%, while that for Nagelkerke is 24.1%, suggesting that more predictors are required to obtain an adequate model.

If autopsy results were available for some of the deaths, it would have been informative to determine the extent to which HIV/AIDS deaths have been misdiagnosed as tuberculosis/other opportunistic infections and vice versa. Where no autopsies are performed, then the reliability of classification of causes of death would vary with the death institution, and available morbidity information about the deceased, among other factors. The higher prevalence of HIV deaths on notifications with multiple causes, and the decrease in HIV prevalence when the multiple causes include tuberculosis, pneumonia, or other opportunistic infections, suggests that some of the deaths attributed to tuberculosis, pneumonia, and other opportunistic infections could be due to HIV disease instead. The magnitude of such misallocation could be estimated if autopsied notifications are available.

Conclusion

The results show that multiple causes of death models lead to significant improvement in predicting HIV-related deaths, as recorded using South African death notifications for 2006 and 2007. The age-gender interaction reveals that gender differences occur mainly in the age group 25–49 years, with females having a much higher likelihood of HIV-related death than males. In cases with multiple causes of death listed, the likelihood of a death being HIV-related when any of tuberculosis, pneumonia, or other opportunistic infection is listed are lower than when none of them is listed. Furthermore, the interaction effects between tuberculosis, pneumonia and other opportunistic infections are significant, and including these terms leads to significant improvement in model performance. This paper points to methods for analyzing multiple causes of death data. Further research is

needed to evaluate the performance of ordinal models for multiple causes of death data, as suggested in the methodology section. These ordinal models could not be evaluated because of data limitations.

Acknowledgment

Disclosure

The authors report no conflicts of interest in this work.

References

1. Weiner L, Bellows MT, McAvoy GH, Cohen EV. Use of multiple causes in the classification of deaths from cardiovascular-renal disease. *Am J Public Health*. 1955;45(4):492–501.
2. Israel RA, Rosenberg HM, Curtin LR. Analytical potential for multiple cause of death data. *Am J Epidemiol*. 1986;124(2):161–179.
3. Santo AH, Pinheiro CE, Jordani MS. Multiple cause of death related to tuberculosis in the state of Sao Paulo, Brazil, 1998. *Revista de Saude Publica*. 2003;37(6):714–721. Portuguese.
4. Mannino DM, Brown C, Giovino GA. Obstructive lung disease deaths in the United States from 1979 through 1993. *Am J Respir Crit Care Med*. 1997;156(3 Pt 1):814–818.
5. Australian Bureau of Statistics. *Multiple Causes of Death Analysis, 1997–2001*. Canberra, Australia: Australian Bureau of Statistics; 2003.
6. Groenewald P, Nannan N, Bourne D, Laubscher R, Bradshaw D. Identifying deaths from AIDS in South Africa. *AIDS*. 2005;19(1): 193–201.
7. Setlhare, K, Forcheh N, Gabaitiri L. Estimating the contribution of HIV/AIDS and related causes to mortality in Botswana. *Eur J Soc Sci*. 2009;9(2):218–230.
8. Bah S. Multiple cause of death statistics in South Africa: Their utility and changing profile over the period 1997 to 2001. PSC Discussion paper series 17(2) Article 1: 2003.
9. Redelings MD, Wise M, Sorvillo F. Using multiple cause-of-death data to investigate associations and causality between conditions listed on the death notification. *Am J Epidemiol*. 2007;166(1):104–108.
10. Wall MM, Huang J, Oswald J, McCullen D. Factors associated with reporting multiple cause of death. *BMC Med Res Methodol*. 2005;5(1):4.
11. Statistics South Africa. *Mortality and Causes of Death from Death Notification, South Africa 2006*. Pretoria, South Africa: Statistics South Africa; 2008.
12. Statistics South Africa. *Mortality and Causes of Death from Death Notification, South Africa 2007*. Pretoria, South Africa: Statistics South Africa; 2009.
13. Anderson BA, Phillips HE. *Adult Mortality (Age 15–64) Based on Death Notification Data in South Africa: 1997–2004*. Pretoria, South Africa: Statistics South Africa; 2006.
14. Hooper WC, Holman RC, Clarke MJ, Chorba TL. Trends in non-Hodgkin lymphoma (NHL) and HIV-associated NHL deaths in the United States. *Am J Hematol*. 2001;66(3):159–166.

Appendix A All causes: HIV with background variables and number of causes listed as predictors

	Chi-square	df	*P* value
Model	23419.723	24	0.000
	Wald		
Death year	20.134	1	0.000
Death institution	662.413	3	0.000
Gender	7.835	1	0.005
Age	2324.199	5	0.000
Education	110.133	4	0.000
Marital status	517.756	4	0.000
Age * gender	75.929	5	0.000
TC	13865.147	1	0.000
Constant	5600.939	1	0.000

Classification table from cross-validation

Predicted HIV disease is one of causes listed

Observed	No	Yes	Percentage correct
No	430412	270	99.9
Yes	29758	98	0.3
Overall percentage			93.5

Abbreviations: TC, total number of causes listed; HIV, human immunodeficiency virus; df, degrees of freedom.

Appendix B One cause listed: HIV with background variables as predictors

	Chi-square	df	*P* value
Model	2010.863	23	0.000
	Wald		
Year of death	70.661	1	0.000
Death institution	188.071	3	0.000
Gender	0.003	1	0.958
Age	412.492	5	0.000
Marital status	127.538	4	0.000
Education	28.483	4	0.000
Age * gender	9.196	5	0.101
Constant	1614.519	1	0.000

Classification table from cross-validation

Predicted HIV disease is one of causes listed

Observed	No	Yes	Percentage correct
No	228368	0	100.0
Yes	4366	0	0.0
Overall percentage			98.1

Abbreviations: HIV, human immunodeficiency virus; df, degrees of freedom.

Appendix C Two causes of death listed: HIV with background variables and tuberculosis as predictors

	Chi-square	df	*P* value
Model	55.4.438	24	0.000
	Wald		
Year of death	5.143	1	0.023
Death institution	437.706	3	0.000
Gender	0.670	1	0.413
Age	1243.552	5	0.000
Education	61.732	4	0.000
Marital status	311.747	4	0.000
Age * gender	38.536	5	0.000
Tuberculosis	200.596	1	0.000
Constant	3373.441	1	0.000

Classification table from cross-validation

Predicted HIV disease is one of causes listed

Observed	No	Yes	Percentage correct
No	134093	0	100.0
Yes	14964	0	0.0
Overall percentage			90.0

Abbreviations: HIV, human immunodeficiency virus; df, degrees of freedom.

Appendix D Two causes of death listed: HIV with background variables and pneumonia as predictors

	Chi-square	df	*P* value
Model	7170.705	24	0.000
	Wald		
Death year	2.273	1	0.132
Death institution	361.423	3	0.000
Gender	1.150	1	0.284
Age	958.606	5	0.000
Education	55.500	4	0.000
Marital status	268.885	4	0.000
Age * gender	32.916	5	0.000
Pneumonia	1540.496	1	0.000
Constant	2770.675	1	0.000

Classification table from cross-validation

Predicted HIV disease is one of causes listed

Observed	No	Yes	Percentage correct
No	134093	0	100.0
Yes	14964	0	0.0
Overall percentage			90.0

Abbreviations: HIV, human immunodeficiency virus; df, degrees of freedom.

Appendix E Two causes of death listed: HIV with background variables and some other opportunistic infection as predictors

	Chi-square	df	P value
Model			
	Wald		
Death year	4.441	1	0.035
Death institution	414.458	3	0.000
Gender	0.307	1	0.580
Age	1076.510	5	0.000
Education	67.439	4	0.000
Age * gender	37.040	5	0.000
Marital status	290.412	4	0.000
Other opportunistic infections	1176.473	1	0.000
Constant	2549.442	1	0.000

Classification table from cross-validation

Predicted HIV disease is one of causes listed

Observed	No	Yes	Percentage correct
No	134093	0	100.0
Yes	14964	0	0.0
Overall percentage			90.0

Abbreviations: HIV, human immunodeficiency virus; df, degrees of freedom.

Appendix F Three causes of death listed: HIV with background variables, tuberculosis, and pneumonia as predictors

	Chi-square	df	P value
Model			
	Wald		
Death year	7.771	1	0.005
Death institution	223.192	3	0.000
Gender	1.788	1	0.181
Age	546.124	5	0.000
Education	22.303	4	0.000
Marital status	95.163	4	0.000
Age * gender	21.736	5	0.001
Tuberculosis	328.972	1	0.000
Pneumonia	729.000	1	0.000
Tuberculosis * pneumonia	113.283	1	0.000
Constant	1128.537	1	0.000

Classification table from cross-validation

Predicted HIV disease is one of causes listed

Observed	No	Yes	Percentage correct
No	50420	0	100.0
Yes	7515	0	0.0
Overall percentage			87.0

Abbreviations: HIV, human immunodeficiency virus; df, degrees of freedom.

Appendix G Three causes of death listed: HIV with background variables, tuberculosis, and some other opportunistic infection as predictors

	Chi-square	df	P value
Model	3412.722	26	0.000
	Wald		
Death year	5.166	1	0.023
Death institution	163.670	3	0.000
Gender	0.609	1	0.435
Age	676.959	5	0.000
Education	33.896	4	0.000
Age * gender	26.640	5	0.000
Marital status	103.275	4	0.000
Tuberculosis	427.017	1	0.000
Other opportunistic infections	729.205	1	0.000
Tuberculosis by other opportunistic infections	277.259	1	0.000
Constant	984.344	1	0.000

Classification table from cross-validation

Predicted HIV disease is one of causes listed

Observed	No	Yes	Percentage correct
No	50420	0	100.0
Yes	7515	0	0.0
Overall percentage			87.0

Abbreviations: HIV, human immunodeficiency virus; df, degrees of freedom.

Appendix H Three causes of death listed: HIV with background variables, pneumonia, and some other opportunistic infection as predictors

	Chi-square	df	P value
Model	3890.916	26	0.000
	Wald		
Year of death	8.280	1	0.004
Death institution	181.424	3	0.000
Gender	1.267	1	0.260
Age	471.188	5	0.000
Education	24.270	4	0.000
Marital status	73.931	4	0.000
Age * gender	23.217	5	0.000
Pneumonia	868.919	1	0.000
Other opportunistic infections	709.644	1	0.000
Pneumonia by other opportunistic infections	62.142	1	0.000
Constant	652.111	1	0.000

Classification table from cross-validation

Predicted HIV disease is one of causes listed

Observed	No	Yes	Percentage correct
No	50420	0	100.0
Yes	7515	0	0.0
Overall percentage			87.0

Abbreviations: HIV, human immunodeficiency virus; df, degrees of freedom.

Appendix I Full model (one of three causes of death listed): HIV with background variables, tuberculosis, pneumonia, and some other opportunistic infection as predictors

	Chi-square	df	P value
Model	3023.156	29	0.000
	Wald		
Death year	8.141	1	0.004
Death institution	128.768	3	0.000
Gender	6.994	1	0.008
Age	170.529	5	0.000
Education	5.608	4	0.230
Age * gender	14.723	5	0.012
Marital status	19.918	4	0.001
Tuberculosis	957.059	1	0.000
Pneumonia	1202.636	1	0.000
Other opportunistic infections	1049.350	1	0.000
Tuberculosis by pneumonia	667.410	1	0.000
Tuberculosis by other opportunistic infections	636.501	1	0.000
Pneumonia by other opportunistic infections	327.720	1	0.000
Constant	113.790	1	0.000

Classification table from cross-validation

Predicted HIV disease is one of causes listed

Observed	No	Yes	Percentage correct
No	17772	29	99.8
Yes	2391	620	20.6
Overall percentage			88.4

Abbreviations: HIV, human immunodeficiency virus; df, degrees of freedom.