

Research

Open Access

## Use of the Social Security Administration Death Master File for ascertainment of mortality status

Enrique F Schisterman<sup>1,2</sup> and Brian W Whitcomb\*<sup>1,2</sup>

Address: <sup>1</sup>Division of Epidemiology, Statistics and Prevention, National Institute of Child Health and Human Development, 9000 Rockville Pike, Bethesda, USA and <sup>2</sup>Department of Nuclear Cardiology, Cedars-Sinai Medical Center, Los Angeles, USA

Email: Enrique F Schisterman - Schistee@mail.nih.gov; Brian W Whitcomb\* - Whitcomb@mail.nih.gov

\* Corresponding author

Published: 05 March 2004

Received: 24 July 2003

*Population Health Metrics* 2004, **2**:2

Accepted: 05 March 2004

This article is available from: <http://www.pophealthmetrics.com/content/2/1/2>

© 2004 Schisterman and Whitcomb; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.

### Abstract

**Objectives:** Internet sources that use the Social Security Administration's (SSA) Death Master File have demonstrated high sensitivity among males for detection of mortality status in comparisons to the National Death Index, but the sensitivity has not been investigated for other demographic groups.

**Methods:** The authors used the SSA Death Master File to determine the mortality status of 374 decedents from the ongoing Patient Outcomes Study at Cedars-Sinai Medical Center whose deaths were confirmed by physicians using hospital records.

**Results:** Decedents identified by the SSA Death Master File were significantly older than those not identified. Foreign-born decedents were significantly less likely to be identified as dead than American-born decedents. Gender and marital status were not significant factors for identification by the SSA Death Master File.

**Conclusion:** The results of this study suggest that Internet sources may be used as an inexpensive and effective tool for determination of mortality status. However, among certain populations use of these databases alone may provide incomplete information.

### Background

Determination of mortality status is an important part of epidemiological studies and many clinical research investigations. Internet sites, such as the Social Security Death Index (SSDI) based on the Social Security Administration (SSA) Death Master File (DMF), are available to researchers for this purpose [1]. The SSA DMF is a database available to the public containing death notices for enrollees in the U.S. Social Security program. This free service is available on the World Wide Web and is updated monthly. The use of databases for ascertainment of mortality status in

epidemiological research is common practice. Many prospective cohort studies evaluate the relation between baseline risk factors and total mortality; by means of linking baseline records with databases the mortality status of study participants can be ascertained. For example, [Gragoudas et al.](#) [2] developed risk score equations to estimate probabilities of death based on an analysis of 2069 patients treated with proton beam radiation for intraocular melanoma and linked to the National Death Index (NDI), a computerized index of death records maintained

by the National Center of Health Statistics for research purposes, and the SSA DMF.

Previous reports have shown the sensitivity of Internet sources for death ascertainment as high as 97.5% among males but as low as 31.1% among females [3] using the NDI as gold standard. The purpose of this paper is to analyze the ability of Internet sites based upon the SSA Death Master File to determine mortality status as a function of gender, ethnic background and additional demographic variables among 374 confirmed decedents.

**Methods**

**Patient population**

For the present study we selected 374 consecutive patients followed up between January 1993 and January 2001 from a population involved in the Myocardial Perfusion Imaging/Patient Outcome (MPI/PO) Study at Cedars-Sinai Medical Center (CSMC), a large hospital in Los Angeles, California, whose deaths occurred at CSMC. Date and cause of death were confirmed by physician review. All demographic information, including name, social security number, place of birth, ethnic group and date of birth, was taken from the hospital admission information.

**Internet sources of vital status**

Internet sites such as <http://Ancestry.com>[1] provide free access to the SSA Death Master File, maintained by the Social Security Administration. The Death Master File contained 65,445,243 records of decedents with social security numbers whose deaths were reported to the SSA and was current through January 2001 at the time of this study. Search tools such as the Social Security Death Index (SSDI) available as a free service on the Internet contain information fields for social security number, surname, given name, date of death, date of birth, last known resi-

dence, location of last benefit, and date and place of issuance. The database is not downloadable, however, software to allow for multiple searches can be easily implemented using packages such as JAVA. Searches can be conducted with any one field or a combination of fields. For this study we used date of birth, social security number and/or first and last name.

Searches were conducted individually and without use of a data matching software package. We considered positive identification for records with exact matches of name, social security number, and dates of death and birth as well as for inexact matches of name with exact match of social security number and/or dates of birth and death [4,5].

**Statistical methods**

Continuous variables are expressed as mean value (standard deviation). The mean differences for continuous variables were compared by t-test (2-tailed). Categorical values are expressed as percentage (standard deviation) and compared using chi-square statistics. Sensitivity and 95% confidence intervals (CI) were estimated. Analysis of variance was performed to estimate adjusted means. Age was divided in four categories based on distribution quartiles. Logistic regression was used to identify the variables that best predict positive detection by the Internet mortality database.

**Results**

The 374 decedents are characterized by identification status in Table 1. The Internet source of the SSA Death Master File identified 330 (88.2%) as dead and failed to identify 44 (11.8%). Those not identified as dead were significantly more likely to be younger and foreign born. We did not observe significant differences in calendar year of birth and of death, gender, or ethnic background.

**Table 1: Comparison of characteristics of study participants from the MPI/PO Study by Identification Status on the SSA DMF.**

|                          | Identified as Dead (n = 330) Mean (SD) | Not identified as Dead (n = 44) Mean (SD) | P-VALUE |
|--------------------------|--|---|---------|
| Age (in years)           | 77.5 (9.9)                             | 70.7 (12.4)                               | <0.01   |
| Year of birth (in years) | 1920.2 (10.0)                          | 1927.5 (12.5) *                           | 0.06    |
| Year of death (in years) | 1997.6 (1.8)                           | 1998.0 (2.3)*                             | 0.13    |
|                          | %                                      | %   |         |
| Male                     | 63.1                                   | 63.6                                      | 1.00    |
| Foreign Born             | 35.8                                   | 59.1                                      | 0.01    |
| African Americans        | 6.7                                    | 13.6                                      | 0.12    |
| Asian                    | 2.4                                    | 2.3                                       | 1.00    |
| Hispanic                 | 3.9                                    | 4.5                                       | 0.69    |

Abbreviations: Myocardial Perfusion Imaging/Patient Outcomes, MPI/PO; Social Security Administration Death Master File, SSA DMF \* This comparison is age adjusted

**Table 2: Sensitivity and 95% confidence intervals (CI) for the SSA DMF in determination of mortality status of decedents from the MPI/PO Study by gender and country of birth.**

|               | Male (n = 239) | 95% CI     | Female (n = 135) | 95% CI     |
|---------------|----------------|------------|------------------|------------|
| American Born | 91.5           | 86.9, 96.1 | 93.8             | 88.5, 99.0 |
| Foreign Born  | 83.7           | 76.4, 91.0 | 77.8             | 65.6, 89.9 |

Abbreviations: confidence interval, CI; Myocardial Perfusion Imaging/Patient Outcomes, MPI/PO; Social Security Administration Death Master File, SSA DMF

The sensitivity of the Internet accessed SSA Death Master File as a function of gender and place of birth is displayed in Table 2. Sensitivity for American-born males, 91.49 percent (95 percent confidence interval: 86.9, 96.1) and females, 93.3 percent (95 percent confidence interval: 88.1, 98.5) is similar. The sensitivity for foreign-born decedents is substantially lower. For foreign-born males it drops to 83.7 percent (95 percent confidence interval: 76.4, 91.0) and for foreign-born females to 77.8 percent (95 percent confidence interval: 65.6, 89.9).

Sensitivity as a function of age is displayed in Table 3. Significant differences in sensitivity were found for age quartiles, ranging from 82.8 percent for ages 41 through 70 (95 percent confidence interval: 75.1, 90.5), to 95.7 percent for ages 85 through 97 (95 percent confidence interval: 91.7, 99.8).

The results of logistic regression are seen in Table 4. Foreign-born decedents had 67% lower odds of being found by the Internet mortality database than American-born counterparts. Decedents in the first age quartile had approximately 85% lower odds of being found than the fourth age quartile. African American decedents had 68% lower odds of being found than were Caucasian decedents ( $p = 0.07$ ). We did not observe statistically significant odds ratios for reaching retirement age (i.e. age 62), gender, or marital status.

**Discussion**

In our study, the internet source of information from the SSA Death Master File demonstrated high and consistent sensitivity for detecting mortality status of both American-born men and women. The sensitivity for American-born decedents was 92.2%, comparable to documented sensitivity for the National Death Index, 87–98% [4,6-8]. However, in foreign-born individuals there is a nearly 10% reduction in sensitivity. The results also suggest that African Americans may have odds as high as 68% of being excluded from Internet databases. Moreover, in our study the odds of sources of the SSA Death Master File finding the youngest decedents were 87% lower than that for the oldest decedents.

The SSA Death Master File is comprised of decedents with social security numbers whose deaths were reported to the Social Security Administration. The SSA reports that in most cases a report of death was made in connection with a claim for Social Security death benefits. In some cases, it is reported to stop Social Security Benefits to the deceased. The primary sources of information utilized for the SSA DMF are relatives of deceased individuals, funeral directors, financial institutions, postal services, as well as other government agencies [9]. Thus the reasons for exclusion from the SSDI include not having a social security number and not having the death reported to the SSA [5,10].

The SSA was originally founded by an act of Congress in 1935 as a retirement program. In 1972 the SSA was required to issue social security numbers (SSNs) to all legally admitted aliens at entry; SSNs are assigned to all persons authorized to work in the US who request them, including newborns. SSNs are required for tax purposes, to get medical coverage or apply for government services. As a result, most Americans and legal aliens have SSNs [11,12].

A recent study compared the SSDI to the NDI using the NDI as the "gold standard" and demonstrated a high sensitivity among men (94.7%), but much lower among women (31.1%) using the first and last name search fields [3]. Our study employed social security number as the primary search field and name as secondary. We found an overall sensitivity of 88.3% for men and 88.1% for women, using confirmed mortality as our "gold standard". We believe one source of this discrepancy to be related to the disproportionate frequency of name changes in women. Having information on social security number has been shown to greatly improve sensitivity, as well as specificity, for sources of mortality [7,9,13,14], possibly by reducing the impact of inexact matches of name (e.g. nicknames, misspelling) [7,15,16]. Investigators using this information have had similar findings among some demographic groups [6,14]. While the Health Insurance Portability and Accountability Act (HIPAA) places certain restrictions on personal information available to researchers, identifiers such as social security number are frequently accessible for studies [17].

**Table 3: Sensitivity and 95% confidence intervals (CI) for the SSA DMF in determination of mortality status of decedents from the MPI/PO Study by quartiles of age at death.**

| AGE                                  | SENSITIVITY | 95 % CI    |
|--------------------------------------|-------------|------------|
| 1 <sup>st</sup> quartile (age 41–70) | 82.8        | 75.1, 90.5 |
| 2 <sup>nd</sup> quartile (age 71–79) | 85.1        | 77.9, 92.3 |
| 3 <sup>rd</sup> quartile (age 80–85) | 87.2        | 80.5, 94.0 |
| 4 <sup>th</sup> quartile (age 86–97) | 95.7        | 91.7, 99.8 |

Abbreviations: confidence interval, CI; Myocardial Perfusion Imaging/Patient Outcomes, MPI/PO; Social Security Administration Death Master File, SSA DMF

**Table 4: Odds Ratios and 95% Confidence Intervals for determination of mortality status of decedents from the MPI/PO Study by the SSA DMF.**

|                              | ODDS RATIO | 95 % CI   | P-VALUE |
|------------------------------|------------|-----------|---------|
| Gender (male)                | 1.1        | 0.5, 2.2  | 0.83    |
| Age 1 <sup>st</sup> Quartile | 1          |           |         |
| 2 <sup>nd</sup> Quartile     | 1.2        | 0.5, 2.6  | 0.72    |
| 3 <sup>rd</sup> Quartile     | 1.5        | 0.6, 3.6  | 0.36    |
| 4 <sup>th</sup> Quartile     | 7.3        | 2.0, 26.8 | <0.01   |
| American born                | 1          |           |         |
| Foreign born                 | 0.3        | 0.1, 0.5  | <0.01   |
| White                        | 1          |           |         |
| African American             | 0.4        | 0.1, 1.1  | 0.07    |
| Asian                        | 3.0        | 0.5, 17.0 | 0.22    |
| Hispanic                     | 3.4        | 0.4, 31.8 | 0.28    |
| Married                      | 1          |           |         |
| Single                       | 3.5        | 0.6, 20.5 | 0.17    |
| Widowed                      | 6.6        | 0.5, 83.8 | 0.15    |
| Divorced                     | 0.6        | 0.2, 1.4  | 0.23    |

Abbreviations: confidence interval, CI; Myocardial Perfusion Imaging/Patient Outcomes, MPI/PO; Social Security Administration Death Master File, SSA DMF

We found that foreign-born decedents had 67% lower odds of being identified by the internet-accessed SSA Death Master File than American-born individuals. A possible explanation for the differential misclassification is related to the eligibility criteria defined by the SSA for receiving death benefits. Foreign nationals and naturalized citizens may have less opportunity to achieve the necessary 40 quarters (10 years) of work in the US to qualify for benefits and thus reduced incentive to report deaths to the SSA. Foreign-born decedents comprised 38% of our study population. The U.S. Census Bureau recently determined that 10.4% of the American population was foreign-born as of March 2000. Immigrant proportions were highest in major urban areas, with Los Angeles, New York City, and San Francisco accountable for the majority of such individuals [18].

Age at death was another determining factor on identification by the SSDI in our study. Older decedents were signif-

icantly more likely to be identified as dead, similar to previous reports [19,20]. In general, as with immigrants who have not had sufficient opportunity to work the necessary 10 years, younger aged decedents are less likely to have achieved qualification for benefits. In this study, the first age quartile ranged from 41–70; it is unlikely to have greatly affected ability to qualify for benefits. We found a significant increase in sensitivity only for decedents older than 85 years at death; sensitivity was approximately 85% for the first 3 quartiles. We also found a significant reduction in sensitivity for determining mortality status of African American decedents. Previous studies have reported difficulties in ascertainment of mortality status in African Americans using databases of such information [7,8,21]. However, these results should be looked at with caution because of the small sample size of African Americans on which they are based.

Our study suggests that the use of the SSDI as the sole source for verification of mortality status might have detrimental effects in research findings if misclassification of mortality status is not accounted for in the analysis. Differential misclassification of mortality status can lead to under/over – estimation of prevalence of outcomes and undesired bias on risk estimates of exposures of interest and their variances. As shown on this paper, this is especially the case, if the exposures of interest are, or related to age, gender, country of birth or race.

Correction methods for bias due to misclassification are available in the literature [22-25]. The matrix method described by Greenland et al. [25] is one alternative to correct odds ratios for misclassification for  $2 \times 2$  tables. Magder et al. [26] showed that when the sensitivity and specificity of a diagnostic test are assumed to be known or can be estimated, this information can be incorporated into the fitting of logistic regression models to estimate risk. They also described an EM algorithm that produces unbiased estimates of the odds ratios and their variances.

This study is limited in its generalizability; the patient population is entirely composed of patients seen for potential heart problems in the Nuclear Cardiology department who agreed to be part of an observational follow-up study. Additionally, while the number of decedents studied is similar to that of similar studies, it is still too low for the analysis of certain subgroups. However, despite the use of the convenience sample for this study, we have no reason to suspect that estimates of overall sensitivity or sensitivity as a function of study variables would be grossly different than population values. Regardless, we encourage application of sensitivity analysis techniques to evaluate different levels of uncertainty with respect to bias. Though we have demonstrated the sensitivity of the SSDI using confirmed decedents, we have not attained similar information for other databases of mortality status, such as the NDI. We have not presented information regarding the specificity of the Internet accessed SSA Death Master File, however, our experience agrees with previous studies that have shown it to be nearly 100% [16,27]. It should be noted that our sole source of demographic information is the hospital admission records. Findings could reflect variance in accuracy as a function of our study variables. However, such information is frequently all that is available to investigators.

## Conclusions

Internet sources provide accurate information for determination of mortality status and may be accessed using the web quickly and inexpensively. The SSA Master Death File from which Internet sources are generated is updated monthly, thus making it particularly useful for researchers

conducting prospective studies with mortality as an endpoint. While gender and marital status have no effect on the sensitivity of SSA Master Death File in our sample, other demographic factors do. There are significant decreases in accuracy among foreign-born decedents, especially women, as well as among African-Americans. For study populations composed largely of these groups, as urban study samples are likely to be, the SSDI may be less effective for determining mortality. Investigators conducting prospective studies should note this as well as the importance of correct information concerning social security number [7,13]. For studies without this information other sources of mortality information should be consulted.

## Abbreviations

NDI, National Death Index; SSA, Social Security Administration; SSDI, Social Security Death Index; WWW, World Wide Web

## Competing interests

None declared.

## Authors' Contributions

ES and BW contributed equally to the preparation of the manuscript.

## References

1. **Social Security Death Index.** [<http://www.ancestry.com/search/rec-type/vital/ssdi/main.htm>] 2003.
2. Gragoudas Evangelos, Li Wenjun, Goitein Michael, Lane Anne Marie, Munzenrider John E., Egan Kathleen M.: **Evidence-Based Estimates of Outcome in Patients Irradiated for Intraocular Melanoma.** *Arch Ophthalmol* 2002, **120**:1665-1671.
3. Sesso HD, Paffenbarger RS, Lee IM: **Comparison of National Death Index and World Wide Web death searches.** *Am J Epidemiol* 2000, **152**:107-11.
4. Porter Pamela Boyer: **Social Security Sleuthing.** Richmond, VA, National Geneological Society; 1999.
5. **Social Security Administration.** 2000, **SSA Publication No. 05-11051.**
6. Boyle CA, Decoufle P: **National sources of vital status information: extent of coverage and possible selectivity in reporting [see comments].** *Am J Epidemiol* 1990, **131**:160-8.
7. Calle EE, Terrell DD: **Utility of the National Death Index for ascertainment of mortality among cancer prevention study II participants.** *Am J Epidemiol* 1993, **137**:235-41.
8. Curb JD, Ford CE, Pressel S, Palmer M, Babcock C, Hawkins CM: **Ascertainment of vital status through the National Death Index and the Social Security Administration.** *Am J Epidemiol* 1985, **121**:754-66.
9. Hill ME, Rosenwaikie I: **The Social Security Administration's Death Master File: the completeness of death reporting at older ages.** *Soc Secur Bull* 2002, **64**:45-51.
10. Porter PB: **Social Security Sleuthing.** Richmond, VA, National Geneological Society; 2003.
11. **Social Security Administration 1.** <http://www.ssa.gov/history/hfaq.html> 2003.
12. **Social Security Administration 2.** <http://www.ssa.gov/history/ssn/ssnchron.html> 2003.
13. Williams BC, Demitrack LB, Fries BE: **The accuracy of the National Death Index when personal identifiers other than Social Security number are used.** *American Journal of Public Health* 1992, **82**:1145-7.
14. Hill ME: **Re: "Comparison of National Death Index and world wide web death searches.** *Am J Epidemiol* 2001, **153**:719.

15. Lash TL, Silliman RA: **A comparison of the National Death Index and Social Security Administration databases to ascertain vital status.** *Epidemiology* 2001, **12**:259-261.
16. Wentworth DN, Neaton JD, Rasmussen WL: **An evaluation of the Social Security Administration master beneficiary record file and the National Death Index in the ascertainment of vital status.** *American Journal of Public Health* 1983, **73**:1270-4.
17. **Department of Health and Human Resources: HIPPA.** <http://www.hhs.gov/ocr/hipaa/> 2003.
18. **Profile of the Foreign-born Population in the United States. US Census Bureau.** 2003.
19. Cowper DC, Kubal JD, Maynard C, Hynes DM: **A primer and comparative review of major US mortality databases.** *Ann Epidemiol* 2002, **12**:462-468.
20. Page WF, Braun MM, Caporaso NE: **Ascertainment of mortality in the US veteran population: World War II veteran twins.** *Mil Med* 1995, **160**:351-355.
21. Acquavella JF, Donaleski D, Hanis NM: **An analysis of mortality follow-up through the National Death Index for a cohort of refinery and petrochemical workers.** *American Journal of Industrial Medicine* 1986, **9**:181-7.
22. Brenner H, Savitz DA, Jockel KH, Greenland S: **Effects of nondifferential exposure misclassification in ecologic studies 8.** *Am J Epidemiol* 1992, **135**:85-95.
23. Copeland KT, Checkoway H, McMichael AJ, Holbrook RH: **Bias due to misclassification in the estimation of relative risk 18.** *Am J Epidemiol* 1977, **105**:488-495.
24. Greenland S: **The effect of misclassification in the presence of covariates 16.** *Am J Epidemiol* 1980, **112**:564-569.
25. Greenland S, Kleinbaum DG: **Correcting for misclassification in two-way tables and matched-pair studies 14.** *Int J Epidemiol* 1983, **12**:93-97.
26. Magner LS, Hughes JP: **Logistic regression when the outcome is measured with uncertainty.** *Am J Epidemiol* 1997, **146**:195-203.
27. Hauser TH, Ho KK: **Accuracy of on-line databases in determining vital status.** *J Clin Epidemiol* 2001, **54**:1267-1270.

Publish with **BioMed Central** and every scientist can read your work free of charge

*"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."*

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

