RESEARCH ARTICLE

Alive or dead: Validity of the Social Security Administration Death Master File after 2011

Matthew A. Levin MD^{1,2} | Hung-Mo Lin ScD³ | Gautham Prabakar BA⁴ |
Patrick J. McCormick MD MEng⁵ | Natalia N. Egorova PhD³

¹Department of Anesthesiology, Perioperative, and Pain Medicine, Icahn School of Medicine at Mount Sinai, New York, New York

²Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, New York

³Department of Population Health Science and Policy, Icahn School of Medicine at Mount Sinai, New York, New York

⁴Paul L. Foster School of Medicine at Texas Tech University Health Sciences Center, El Paso, Texas

⁵Department of Anesthesiology & Critical Care Medicine, Memorial Sloan Kettering Cancer Center, New York, New York

Correspondence

Matthew A. Levin, MD, Department of Anesthesiology, Perioperative, and Pain Medicine, Icahn School of Medicine at Mount Sinai, New York, NY.
Email: matthew.levin@mssm.edu

Funding information

National Cancer Institute of the National Institutes of Health, Grant/Award Number: R25CA020449

Objective: To determine the reliability of the Social Security Death Master File (DMF) after the November 2011 changes limiting the inclusion of state records.

Data Sources: Secondary data from the DMF, New York State (NYS) and New Jersey (NJ) Vital Statistics (VS), and institutional data warehouse.

Study Design: Retrospective study. Two cohorts: discharge date before November 1, 2011, (pre-2011) or after (post-2011). Death in-hospital used as gold standard. NYS VS used for out-of-hospital death. Sensitivity, specificity, Cohen's Kappa, and 1-year survival calculated.

Data Collection Methods: Patients matched to DMF using Social Security Number, or date of birth and Soundex algorithm. Patients matched to NY and NJ VS using probabilistic linking.

Principal Findings: 97 069 patients January 2007-March 2016: 39 075 pre-2011; 57 994 post-2011. 3777 (3.9 percent) died in-hospital. DMF sensitivity for in-hospital death 88.9 percent ($\kappa = 0.93$) pre-2011 vs 14.8 percent ($\kappa = 0.25$) post-2011. DMF sensitivity for NY deaths 74.6 percent ($\kappa = 0.71$) pre-2011 vs 26.6 percent ($\kappa = 0.33$) post-2011. DMF sensitivity for NJ deaths 62.6 percent ($\kappa = 0.64$) pre-2011 vs 10.8 percent ($\kappa = 0.15$) post-2011. DMF sensitivity for out-of-hospital death 71.4 percent pre-2011 ($\kappa = 0.58$) vs 28.9 percent post-2011 ($\kappa = 0.34$). Post-2011, 1-year survival using DMF data was overestimated at 95.8 percent, vs 86.1 percent using NYS VS.

Conclusions: The DMF is no longer a reliable source of death data. Researchers using the DMF may underestimate mortality.

KEYWORDS

death index, federal policy, mortality, patient outcomes

1 | INTRODUCTION

Large retrospective research studies often draw on multiple secondary sources such as electronic health records and claims data from private insurers. In-hospital mortality is available in claims data but to get mortality outside of the hospital, researchers link claims data with vital statistics records. Acquiring these data can be a difficult and time-consuming process, particularly if the patient died

at a different health care facility or at home. In the United States, specifically, it is difficult to determine mortality status due to multiple separate health and vital statistic databases at the local, state, and federal levels. Mortality measurement of very elderly patients may also be difficult.¹ Researchers and clinicians have thus used the Social Security Death Master File (DMF—also known as the “death index”) as an easily accessible, fairly inexpensive, and reliable source of mortality data.^{2–5} The DMF has a relatively short time lag

in reporting death of only 4 to 6 months.⁶ The master file includes data on both patients who died at home and those who died within a health care facility. The DMF is made available via a secure website, with a variety of access options.⁷ The DMF contains mortality data for over 86 million deceased individuals since 1936 created from Social Security Administration (SSA) records.

On November 1, 2011, the SSA made rigorous and strict changes to the DMF in response to fears of identity theft and fraud.⁸ These changes were made to better comply with Section 205(r) of the Social Security Act, added by the Act of April 20, 1983, [Pub L No. 98-21, 97 Stat. 65, 130] which prohibited the SSA from disclosing the state death records SSA receives through its contracts with the states, except in limited circumstances.⁹ These changes resulted in the removal of approximately 4 million historical records, and the addition of approximately 1 million fewer records of deceased per year.¹⁰ The reduced DMF is officially referred to as the Limited Access Death Master File (LADMF, henceforth referred to simply as DMF). The potential impact of this change on the utility of the DMF has been as source of concern for several years; however, to our knowledge, it has not been rigorously investigated.⁶

We hypothesized that the 2011 change would greatly reduce the validity of the DMF as a source of mortality data. We sought to analyze the effect of the 2011 DMF change by comparing DMF data to a gold standard of in-hospital mortality. As a secondary endpoint, we also compared DMF data to in-hospital and out-of-hospital mortality data from the New York State Statewide Planning and Research Cooperative System (SPARCS) discharge database linked with the New York State Vital Statistics (NYS VS) death records and to mortality data from the New Jersey Vital Statistics (NJ VS) records. Finally, to evaluate the potential impact of the change in DMF policy on estimation of long-term survival, we modeled 1-year survival curves using the DMF data pre- and post-2011 and compared the results to 1-year survival curves generated using the SPARCS/NYS VS mortality data.

2 | METHODS

Institutional Review Board approval from our institution (Icahn School of Medicine at Mount Sinai), the New York State Privacy Board, and the New Jersey Department of Health was obtained prior to beginning this study.¹¹ We retrospectively queried our perioperative data warehouse for all anesthesia/surgery cases from January 1, 2007, to March 31, 2016. Medical admissions were not included in the cohort. Inclusion criteria were age >18, American Society of Anesthesiologists (ASA) Physical Status 3 to 6. The ASA Physical Status score is a measure of comorbidity and overall health, where an ASA 1 patient is a normal healthy patient, an ASA 5 is a "moribund patient who is not expected to survive without the operation," and an ASA 6 is an organ donor.¹² Healthier patients (ASA 1,2) were excluded in order to enhance the mortality rate in the cohort. For patients with multiple anesthetics/hospital admissions, only the most recent was preserved. Hospital date of death, discharge date,

and discharge status were obtained from our institutional data warehouse. If a patient had multiple anesthetics during the study period, only the most recent anesthetic was considered.

We a priori excluded all cases in a four-month window surrounding November 1, 2011, (ie, September 1, 2011, to December 31, 2011), in order to reduce confusion arising from cases where the date of discharge was before November 1, 2011, but the date of death was after November 1, 2011. A four-month window was chosen based on the clinical observation that most patients who die in hospital do so within 2 months of their procedure.

2.1 | Determination of the gold standard of death in hospital

A patient was considered to have died in hospital if the hospital date of death was less than or equal to their discharge date, and the discharge status was one of "expired," "patient has expired," or "organ harvest." All other discharge statuses indicated that the patient left the hospital alive. If the date of death was null and the discharge status indicated death, then date of death was set to the date of discharge, and the patient was also considered to have died in hospital. It is important to note that the absence of hospital death date is *not* a gold standard for "not dead," that is, the absence of a hospital death date is not a guarantee that the patient is still alive after discharge. In fact, identifying death after discharge is one of the main motivations of using the DMF.

2.2 | Determination of DMF date of death

The SSA makes the DMF available to authorized users as a base file, supplemented by either weekly or monthly updates. The publicly available DMF is distributed through the National Technical Information Service (NTIS) of the Department of Commerce.¹³ Our institutional subscription to the DMF began in the third quarter of 2011, with a base file containing approximately 90 million records current through August 31, 2011. We have kept our local DMF file current by applying monthly updates as mandated by the subscriber agreement.¹⁴ The monthly update is automated using a locally developed program that is freely available.¹⁵ Base files provided after the November 1, 2011, change had approximately 4 million records removed.¹⁰ However, based on our review of the historical monthly raw data files, the monthly updates since that time have never included a bulk deletion. Thus, we were in a unique position of being able to compare the historical coverage and validity of the DMF with the current post-2011, accuracy.

Patients were matched to our local copy of the DMF using the Social Security Number (SSN). The match was run on October 16, 2016. Typically, there is a 6- to 8-week delay before a death will appear in the DMF. Prior to matching, we validated SSN's as per the published SSA description.¹⁶ Additional matching on name and date of birth was performed by matching on date of birth plus the first three letters of the first and last name, or date of birth and the Soundex algorithm on the full first and last name.¹⁷

2.3 | Determination of NYS date of death

We used New York State Vital Statistics death records linked with mandatory all payer New York State discharge abstract data from the Statewide Planning and Research Cooperative System (SPARCS) for the years 2007-2014. We used the identifiable version of the data after obtaining approval by NYS Department of Health Data Governance Committee.¹⁸ There is a delay in the release of SPARCS data so that the dataset was limited to discharges on or before December 31, 2014, and deaths on or before December 31, 2015. The New York State Department of Health and Human Services linked the data using a proprietary and non-public probabilistic linking methodology. The linked dataset included date and cause of death, unique patient identifier, hospital name and unique facility identification number as well as admission and discharge dates, patient sex, date of births, and residency, primary and up to 24 secondary diagnoses, primary and up to 14 secondary procedures. The date of death for patients who underwent surgical procedure at our institution was identified through a discharge disposition of "deceased" or by using the date of death from the New York State vital statistics death records. Unlike our institutional data, "Organ harvest" is not included as a discharge disposition.

2.4 | Definition of NYS out-of-hospital deaths

Participation in SPARCS is mandatory; therefore, SPARCS contains all hospital discharges. Each discharge is linked with the NY State Vital Statistics data and contains both in-hospital and out-of-hospital deaths that occurred within 1-year postdischarge, for all New York State residents. We defined an out-of-hospital death as a NY State resident who was treated at our institution, died in NY, but did not die in our institution (ie, they may have died in another health care facility/hospital, but not in our hospital). Out-of-hospital death data were available through December 31, 2015.

2.5 | Determination of NJ date of death

The New Jersey Department of Health Office of Vital Statistics and Registry maintains death records for all deaths in New Jersey from 1918 to the present.¹⁹ Data from our cohort were linked to the New Jersey Vital Statistics data using "The Link King" SAS plugin (version 9.0).²⁰ The last date of death available in the NJ file was December 31, 2015; therefore, analysis was limited to patients who died on or before that date.

2.6 | Statistical analysis

Data were split into two cohorts based on discharge date: discharge before November 1, 2011 (pre-2011), and discharge after November 1, 2011 (post-2011). Chi-square analysis was used to compare the percentage of cases where a DMF death was found for a known hospital death and the sensitivity (positive percent agreement) and specificity (negative percent agreement) of the DMF for in-hospital

death were calculated using the following definitions. A *true positive* was a DMF record found for a patient who died in hospital, while a *false positive* was a DMF record found for a patient who did *not* die in hospital. A *true negative* was no DMF record for a patient who did *not* die in hospital, and a *false negative* was no DMF record for a patient who died in hospital. All false positives were manually checked via chart review. In addition, Cohen's kappa coefficient was calculated.²¹ The kappa coefficient is a preferred measure of the degree of agreement between two independent measures.²² The same analysis (chi-square, sensitivity/specificity, and Kappa) was performed for NYS VS death vs in-hospital death, for DMF vs NYS VS, for out-of-hospital death, and for DMF vs NJ death.

One-year Kaplan-Meier survival curves were generated from the DMF and SPARCS mortality data. This would simulate a hypothetical study with mortality as a primary endpoint. The curves are presented by cohort of patients discharged before and after November 2011 to emphasize the effect DMF change. The analysis was limited to New York State residents only with discharges on or before December 31, 2014. Survival data were censored at 1 year after the date of surgery for cases who survived beyond 1 year or were missing death data. Statistical analysis was performed using R 3.4.1 (R Foundation, Vienna, Austria).

3 | RESULTS

A total of 101 010 patients met inclusion criteria. After exclusion of patients with discharge date in the four-month exclusion window ($n = 3941$), the final dataset contained 97 069 patients: 39 075 with a discharge date prior to November 2011, and 57 994 with a discharge date after November 2011. Demographics are shown in Table 1. While the majority of patients were from New York State (83 869, 86.4 percent), there were 9046 (9.3 percent) patients from NJ and the cohort contained patients from all 50 states. The cohorts were well matched in terms of age, gender, and ASA Physical Status. The percentage of SSN's available in the post-2011 cohort was significantly lower than in the pre-2011 cohort (76.3 percent vs 87.3 percent, $P < 0.001$), and thus, the number of DMF records matched on name and date of birth was correspondingly higher (17.8 percent vs 4.1 percent, $P < 0.001$). The overall in-hospital mortality significantly dropped in the post-2011 period (5.4 percent vs 2.9 percent, $P < 0.001$). The overall number of deaths found in the DMF for the pre-2011 cohort was 10 396, whereas in the post-2011 cohort (a larger cohort), the number of death records found in the DMF was significantly fewer, only 2608 ($P < 0.001$). Figure 1 top panel illustrates the drop in the number of death records found in the DMF pre-2011 vs post-2011.

3.1 | Validity of the DMF for in-hospital death

Overall 3777 patients died in hospital, 2102 (5.4 percent) pre-2011 and 1675 (2.9 percent) post-2011. 88.9 percent ($n = 1868$) of patients who died in hospital pre-2011 had a death record in the

TABLE 1 Demographics of entire cohort

Variables	All	January 2007-October 2011	November 2011-March 2016	P-value
N	97 069	39 075	57 994	
Age, y	61.7 (16.4)	61.9 (16.5)	61.5 (16.3)	<0.001
Age group, y				
≤45	15 064 (15.5%)	6077 (15.6%)	8987 (15.5%)	<0.001
46-54	14 477 (14.9%)	5920 (15.2%)	8557 (14.8%)	
55-64	21 990 (22.7%)	8670 (22.2%)	13 320 (23.0%)	
65-74	22 832 (23.5%)	8680 (22.2%)	14 152 (24.4%)	
75-84	16 357 (16.9%)	7049 (18.0%)	9308 (16.0%)	
≥85	6349 (6.5%)	2679 (6.9%)	3670 (6.3%)	
Gender				
Female	48 298 (49.8%)	19 585 (50.1%)	28 713 (49.5%)	0.063
Male	48 771 (50.2%)	19 490 (49.9%)	29 281 (50.5%)	
Race				
African American (Black)	16 022 (16.5%)	5916 (15.1%)	10 106 (17.4%)	<0.001
Asian	4331 (4.5%)	1700 (4.4%)	2631 (4.5%)	
Caucasian (White)	54 410 (56.1%)	26 175 (67.0%)	28 235 (48.7%)	
Other	16 742 (17.2%)	3192 (8.2%)	13 550 (23.4%)	
Unknown	5564 (5.7%)	2092 (5.4%)	3472 (6.0%)	
Payor				
Commercial or Private	29 984 (30.9%)	12 998 (33.3%)	16 986 (29.3%)	<0.001
Medicaid	15 909 (16.4%)	5504 (14.1%)	10 405 (17.9%)	
Medicare	47 360 (48.8%)	19 223 (49.2%)	28 137 (48.5%)	
Other	3180 (3.3%)	1058 (2.7%)	2122 (3.7%)	
Self Pay	636 (0.7%)	292 (0.7%)	344 (0.6%)	
ASA physical status				
3	73 963 (76.2%)	29 803 (76.3%)	44 160 (76.1%)	<0.001
4	21 609 (22.3%)	8767 (22.4%)	12 842 (22.1%)	
5	1467 (1.5%)	487 (1.2%)	980 (1.7%)	
6	30 (0.0%)	18 (0.0%)	12 (0.0%)	
NYS Resident	83 869 (86.4%)	32 855 (84.1%)	51 014 (88.0%)	<0.001
NJ Resident	9046 (9.3%)	4258 (10.9%)	4788 (8.3%)	<0.001
Other states	4154 (4.3%)	1962 (5.0%)	2192 (3.8%)	<0.001
SSN available and valid	78 350 (80.7%)	34 110 (87.3%)	44 240 (76.3%)	<0.001
In-hospital Death	3777 (3.9%)	2102 (5.4%)	1675 (2.9%)	<0.001
Deaths in DMF	13 004 (13.4%)	10 396 (26.6%)	2608 (4.5%)	<0.001
SSN match	12 111 (93.1%)	9967 (95.9%)	2144 (82.2%)	<0.001
Name/date of birth match	893 (6.9%)	429 (4.1%)	464 (17.8%)	<0.001

ASA, American Society of Anesthesiologists; DMF, Death Master File; NJ, New Jersey; NYS, New York State; SSN, Social Security Number.

DMF but only 14.8 percent ($n = 248$) post-2011, $P < 0.001$. There were 18 false positives in the pre-2011 cohort (DMF death record date before date of discharge but no recorded death in hospital). Of

them, 17 were matched directly on SSN; one case was matched on name. The majority of false positives (12/18, 66 percent) were due to no hospital death date recorded and a discharge disposition other

FIGURE 1 Top Panel: Decrease in number of death records found in DMF After 2011; Bottom Panel: Comparison of death records found in DMF vs NYS VS [Color figure can be viewed at wileyonlinelibrary.com]

than “expired,” “patient has expired,” or “organ harvest.” In addition, two were due to the hospital date of death being recorded as after the date of discharge (2 days and 5 days afterward). The remaining four had a DMF death date on or before the date of discharge but

a discharge disposition of “Home Health Care” or “Skilled Nursing Facility.” Sensitivity of the DMF pre-2011 was 88.87 percent, specificity was 99.95 percent, and the Kappa coefficient was 0.93. (Table 2) In the post-2011 period, there were six false positives, all

TABLE 2 Validity of DMF and NYS VS for in-hospital death

Validity of DMF for In-hospital death			
	January 2007-October 2011	November 2011-March 2016	P-value
In-hospital death	n = 2102	n = 1675	
DMF In-hospital death	1868 (88.9%)	248 (14.8%)	<0.001
DMF sensitivity	88.87%	14.81%	—
DMF specificity	99.95%	99.99%	—
DMF Kappa coefficient	0.93	0.25	—
Validity of NYS Vital Statistics for In-hospital death			
	January 2007-October 2011	November 2011-December 2015	P-value
In-hospital death	n = 2089	n = 1181	
NYS VS In-hospital death	2085 (99.8%)	1179 (99.8%)	1.00
NYS VS Sensitivity, In-hospital	99.81%	99.83%	—
NYS VS Specificity, In-hospital	100.00%	99.99%	—
NYS VS Kappa coefficient, overall	0.999	0.998	—

DMF, Death Master File; NYS VS, New York State Vital Statistics.

TABLE 3 Validity of DMF vs NYS Vital Statistics and NJ Vital Statistics and validity of DMF for out-of-hospital death

Validity of DMF vs NYS Vital Statistics			
	January 2007-October 2011	November 2011-December 2015	P-value
NYS VS all deaths	n = 10 267	n = 5365	
DMF deaths	7662 (74.6%)	1426 (26.6%)	<0.001
DMF sensitivity	74.63%	26.58%	—
DMF specificity	94.19%	97.92%	—
DMF Kappa coefficient	0.71	0.33	—
Validity of DMF vs NJ Vital Statistics			
	January 2007-October 2011	November 2011-December 2015	P-value
NJ VS deaths	n = 1167	n = 549	P-value
DMF deaths	731 (62.6%)	59 (10.7%)	<0.001
DMF sensitivity	62.64%	10.75%	—
DMF specificity	95.86%	99.13%	—
DMF Kappa coefficient	0.64	0.15	—
Validity of DMF for Out-of-hospital death, NYS residents only			
	January 2007-October 2011	November 2011-December 2015	P-value
	n = 8406	n = 4287	
NYS VS OOH deaths	6003 (71.4%)	1241 (28.9%)	<0.001
DMF sensitivity OOH	71.41%	28.95%	—
DMF specificity OOH	87.68%	97.33%	—
DMF Kappa coefficient OOH	0.58	0.34	—

Notes: NJ deaths through December 31, 2015, NYS deaths through December 31, 2015. OOH death defined as NY resident who was treated at our institution, died in NY, but did not die in our institution, through December 31, 2015.

DMF, Death Master File; NJ VS, New Jersey Vital Statistics; NYS VS, New York State Vital Statistics; OOH, out of hospital.

of which were due to a hospital disposition other than “expired,” “patient has expired,” or “organ harvest.” For all six cases, the DMF date of death matched hospital discharge date. The sensitivity of

the DMF dropped to 14.81 percent while the specificity remained high at 99.99 percent. The Kappa coefficient decreased to 0.25 (Table 2).

FIGURE 2 Kaplan-Meier 1-year survival curves for pre-2011 (top panel) and post-2011 (bottom panel). DMF shown in blue, and New York State vital statistics death records (NYS) in red. Post-2011 the curves diverge and the DMF overestimates survival [Color figure can be viewed at wileyonlinelibrary.com]

When analyzed by various subgroups, the validity of the DMF post-2011 was highest for Medicare patients ($\kappa = 0.3$), older patients ($\kappa = 0.33$ for age 75-84; $\kappa = 0.32$ for age ≥ 85), female patients ($\kappa = 0.29$), Asian patients ($\kappa = 0.31$), and ASA PS 3 patients ($\kappa = 0.29$). In general, it remained low for all subgroups (see Tables S1-S5).

3.2 | Validity of NYS VS for in-hospital death

There were 77 412 patients discharged on or before December 31, 2014; 39 126 pre-2011 and 38 286 post-2011. The overall match rate to NYS VS was high, 98.2 percent pre-2011 and 96.1 percent post-2011. There was one false positive pre-2011 and three false positives post-2011. The sensitivity of SPARCS for in-hospital death did not significantly change over the entire period, 99.8 percent pre-2011 vs 99.8 percent post-2011 (Table 2). The Kappa coefficient likewise remained very high at 0.999 pre-2011 vs 0.998 post-2011.

3.3 | Validity of DMF vs NYS VS

Comparing the DMF to SPARCS linked to NYS VS (Table 3), 75.1 percent of NYS VS deaths were found in the DMF pre-2011, vs only 26.1 percent of deaths after 2011. Figure 1 bottom panel illustrates the drop off in DMF data vs NYS VS per year. The sensitivity of the DMF for capturing NYS VS deaths dropped from 74.6 percent pre-2011 to 26.6 percent post-2011, and the Kappa coefficient decreased from 0.72 to 0.33.

3.4 | Validity of DMF for NYS out-of-hospital death

There were 12 693 out-of-hospital deaths found in the NYS VS records on or before December 31, 2015, 8406 pre-2011 and 4287 post-2011. Pre-2011, the sensitivity of the DMF for out-of-hospital deaths was 71.4 percent, and kappa was 0.58 (Table 3). Post-2011, the sensitivity dropped to 28.95 percent and the kappa decreased to 0.34.

3.5 | Validity of DMF vs NJ VS

Mortality data were available for NJ patients through December 31, 2015. There were 1716 deaths, 1167 pre-2011 and 549 post-2011. Sensitivity of the DMF for NJ deaths declined from 62.6 percent ($\kappa = 0.64$) pre-2011 to 10.8 percent ($\kappa = 0.15$) post-2011 (Table 3).

3.6 | Effect of DMF change on a study using 1-year mortality as an endpoint

As an illustration of the impact the change in the DMF would have on clinical research, we modeled a study with 1-year mortality as an endpoint. We generated Kaplan-Meier survival curves for the pre-2011 and post-2011 cohorts, using mortality data from the DMF and from SPARCS/NYS VS. The results are shown in Figure 2A,B. Pre-2011, the 1-year survival computed using the DMF vs SPARCS/NYS VS was similar, 83.3 percent vs 82.4 percent, respectively. If

the same study had been performed after 2011, the 1-year survival using DMF data would now appear to 95.8 percent, vs 86.1 percent when using SPARCS/NYS VS. In other words, the 1-year mortality rate would be underestimated by about 10 percent if using DMF death records.

4 | DISCUSSION

This retrospective, single-center study of a large surgical cohort found that the Social Security Death Master File is no longer a reliable single source for mortality data. This was shown to be true for both in-hospital deaths at our institution, as well as out-of-hospital deaths in NY state, and overall deaths among NJ residents. Historically, the DMF was shown to be a reliable source of death data for elderly individuals, with 96 percent of deaths of patients over the age of 65 captured.²³ The 2011 change restricting state contributions to the public file has dramatically reduced its coverage. The implications of this change have been a source of concern since at least 2012.^{6,24-26} Peterss et al²⁷ recently demonstrated that the DMF performs poorly relative to a locally maintained clinical database, but the number of patients and deaths included in the study was small and limited to patients with aortic disease. Ashley et al²⁸ used life claims data from a large insurer to show that even prior to 2011, only 75 percent of deaths appeared in the DMF. This study was also of a relatively small cohort ($n = 693$). The current study is the first to our knowledge to examine the effects of the change in the DMF using a very large cohort of general and specialty surgical cases. The clinical endpoint of death in-hospital was used as a reliable gold standard for mortality.

In 2013, limitations on DMF access were added as part of the Bipartisan Budget Act of 2013.²⁹ Researchers wishing DMF data must be certified by the NTIS. Certification requires that a researcher demonstrates a legitimate purpose to access DMF data, and has systems and procedures to secure the data. This extra layer of requirements creates more barriers to research use of the DMF.

The lack of an inexpensive and broad-coverage source of vital status makes it difficult for researchers and their institutions to determine the true time to death for patients who transition care to another facility. Lack of data results in reported survival that is artificially higher and reduces the number of true deaths in any study. With fewer deaths, effect size is smaller, and the power of a study to find a true difference between groups is reduced.

The DMF is used not only for research but also by hospitals for quality, safety, and performance monitoring.⁶ The reduced coverage of the limited DMF will affect these efforts as well.

While there are alternative sources of death data for patients in the United States, few offer the broad coverage of the DMF. State and local vital records may be difficult or impossible to obtain for researchers in different localities. Commercial obituary record agencies such as ObituaryData.com, while comprehensive, are expensive.³⁰ ObituaryData.com data are also limited to the year 2000 onward. The Centers for Disease Control National Death Index (NDI) is a centralized database of death records that can be used

for research studies, but not for administrative tasks.³¹ Even for research studies, there is no method to automate the submission of data requests to the CDC, nor permission to maintain a local copy of death records for shared use by researchers. Requests are expensive, must be submitted via mail on a compact disk, and have a long turnaround time.

Other US death data sources are restricted to subpopulations. For example, the US Veterans Administration (VA) offers a Vital Status File that is used to determine mortality for subjects in VA research trials, and also a Beneficiary Identification Records Locator Subsystem Death File which covers veterans and their survivors who received benefits from the VA. The Centers for Medicare and Medicaid Services (CMS) includes vital status in data extracts, but these extracts cannot be used to query for a specific patient.

In contrast to the DMF, local state databases likely remain a reliable source of death data. The NYS SPARCS data linked with NYS Vital Statistics death records showed a high degree of agreement with our gold standard of in-hospital death throughout the study period and were unaffected by the change to the DMF. State vital death records accurately capture deaths that occurred in hospital (gold standard—Table 3), and thus, there should be no concern that they capture death outside of the hospital inaccurately. The limitation of state vital death records could be that they may not correctly capture deaths for non-residents and for residents who die out of state. There is often also a significant lag time for release of death date. The NYS SPARCS/Vital Statistics data that we used has at least a one-year delay in the reporting of mortality data, as did the NJ Vital Statistics data.

4.1 | Decline in SSNs available

An interesting finding in our study was that the number of available SSNs in our dataset declined by about 15 percent between the two periods. The Social Security Number has become a fixture of American life since the first cards were issued in 1935. The near universality of the number among Americans adults has led the SSN to become a de facto national identification number. However, concern over the SSN's use in varied contexts and the rise of identity theft have created a strong push at governmental and commercial levels to eliminate the use of SSNs as identifiers. The Medicare Access and CHIP Reauthorization Act (MACRA) of 2015 requires the US Centers for Medicare & Medicaid Services to remove SSNs from all Medicare cards. The current Medicare identifier is based on the patient's SSN and will be replaced with a new identifier starting in April 2018.³² Hospitals have fewer reasons to require an SSN from a patient since payors frequently do not require the SSN in order to submit an insurance claim.

4.2 | Limitations

Our analysis was limited to data from a single tertiary care center, and the majority of the patients were residents of a single state (New York). It is possible that other states might have more permissive

death data use agreements with the SSA and that DMF data would be more reliable for those states. This is unlikely given that the changes made to the DMF were done at the federal level. Indeed, analysis of the validity of DMF vs the New Jersey Vital Statistics data showed a similar decline to New York State. Another possible limitation is that the in-hospital mortality rate declined by almost 50 percent between the two periods from 5.4 percent to 2.9 percent. We are unable to speculate on the source of this decline since we did not examine cause of death or severity of illness. It is possible the reduced DMF match rate was simply the result of fewer deaths in the post-2011 period. The decline in DMF death records found however was disproportionately greater than the decline in in-hospital mortality.

Recently, the NTIS announced that in fiscal year 2018, the SSA will add over 8 million death records to the full and publicly available DMF over the course of several months (personal communication from NTIS to author MAL). While this will no doubt increase the coverage of the DMF, the full impact of the addition of these records remains to be seen.

5 | CONCLUSION

The utility of the DMF as a reliable source of death data has declined significantly as a result of the changes made by the SSA in November 2011. Although there is now an effort underway to increase the coverage of the DMF, researchers relying on the DMF will significantly underestimate mortality in outcome and population studies. A more reliable source of mortality data must be found to continue performing high-quality outcome studies in the United States.

ACKNOWLEDGMENTS

Joint Acknowledgment/Disclosure Statement: Dr. McCormick acknowledges that research reported in this publication was supported by the National Cancer Institute of the National Institutes of Health under Award Number R25CA020449. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The remaining authors acknowledge that they received no financial or material support for this project. There are no other disclosures.

ORCID

Matthew A. Levin <https://orcid.org/0000-0002-6013-2684>

Hung-Mo Lin <http://orcid.org/0000-0003-3176-6570>

Patrick J. McCormick <http://orcid.org/0000-0002-7062-2409>

REFERENCES

- Leonid A, Gavrilov NSG. Mortality measurement at advanced ages: a study of the Social Security Administration Death Master File. *N Am Actuar J*. 2011;15(3):432.

2. Hauser TH, Ho KK. Accuracy of on-line databases in determining vital status. *J Clin Epidemiol*. 2001;54(12):1267-1270.
3. Huntington JT, Butterfield M, Fisher J, Torrent D, Bloomston M. The Social Security Death Index (SSDI) most accurately reflects true survival for older oncology patients. *Am J Cancer Res*. 2013;3(5):518.
4. Hanna DB, Pfeiffer MR, Sackoff JE, Selik RM. Comparing the national death index and the social security administration's death master file to ascertain death in HIV surveillance. *Public Health*. 2009;124:850-860.
5. Min-Woong S, Noreen A, Charles M, Hynes DM. Accuracy and completeness of mortality data in the Department of Veterans Affairs. *Popul Health Metr*. 2006;4:2.
6. da Graca B, Filardo G, Nicewander D. Consequences for healthcare quality and research of the exclusion of records from the Death Master File. *Circ Cardiovasc Qual Outcomes*. 2013;6(1):124-128.
7. Social Security Death Master File. <https://www.ssdmf.com/>. Accessed September 18, 2018.
8. General O of the I. Hearing on Social Security's Death Records|Office of the Inspector General, SSA. <https://oig.ssa.gov/newsroom/congressional-testimony/hearing-social-securitys-death-records>. Accessed September 18, 2018.
9. Social Security Act: Evidence, Procedure, and Certification for Payment. United States of America: United States Congress: Sec. 205. [42 U.S.C. 405]. https://www.ssa.gov/OP_Home/ssact/title02/0205.htm. Accessed September 18, 2018.
10. National Technical Information Service. IMPORTANT NOTICE: Change in Public Death Master File Records. <https://classic.ntis.gov/assets/pdf/import-change-dmf.pdf>. Accessed September 18, 2018.
11. NJDOH Data Requirements related to IRB Submissions. <https://sites.rowan.edu/officeofresearch/compliance/irb/njdoh/>. Accessed September 18, 2018.
12. ASA Physical Status Classification System. 2014. <http://www.asahq.org/resources/clinical-information/asa-physical-status-classification-system>. Accessed September 18, 2018.
13. National Technical Information Service. NTIS: Death Master File Download. <https://dmf.ntis.gov/>. Accessed September 18, 2018.
14. National Technical Information Service. NTIS: Mandatory Requirements. <https://dmf.ntis.gov/requirements.html>. Accessed September 18, 2018.
15. Levin MA. DMF Loader. <https://github.com/levin-lab/dmf-loader>. Accessed September 18, 2018.
16. Social Security Administration. SSA - POMS: RM 10201.035 - Invalid Social Security Numbers (SSNs). 2011. <https://secure.ssa.gov/apps10/poms.nsf/lnx/0110201035>. Accessed September 18, 2018.
17. National Archives. Soundex System|The Soundex Indexing System. 2007. <https://www.archives.gov/research/census/soundex.html>. Accessed September 18, 2018.
18. New York State Department of Health. Data Access - Statewide Planning and Research Cooperative System. 2018. <https://www.health.ny.gov/statistics/sparcs/access/>. Accessed September 18, 2018.
19. State of New Jersey Department of Health. Department of Health| Vital Statistics. <https://www.state.nj.us/health/vital/>. Accessed September 18, 2018.
20. The Link King Record Linkage and Consolidation Software. <http://www.the-link-king.com/index.html>. Accessed September 18, 2018.
21. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas*. 1960;20:37-46.
22. Watson PF, Petrie A. Method agreement analysis: A review of correct methodology. *Theriogenology*. 2010;73(9):1167-1179.
23. Hill ME, Rosenwaite I. The Social Security Administration's Death Master File: the completeness of death reporting at older ages. *Soc Sec Bull*. 2001;64(1):45-51.
24. Blackstone EH. Demise of a vital resource. *J Thorac Cardiovasc Surg*. 2012;143(1):37-38.
25. Sack K. Researchers Wring Hands as U.S. Clamps Down on Death Record Access. 2012. <http://www.nytimes.com/2012/10/09/us/social-security-death-record-limits-hinder-researchers.html>. Accessed September 18, 2018.
26. Winn D. Could the recent change in availability of death master file data affect your research? <https://epi.grants.cancer.gov/blog/archive/2012/05-24.html>. Accessed September 18, 2018.
27. Peterss S, Charilaou P, Ziganshin BA, Eleftheriades JA. Assessment of survival in retrospective studies: The Social Security Death Index is not adequate for estimation. *J Thorac Cardiovasc Surg*. 2017;153(4):899-901.
28. Ashley T, Cheung L, Wokanovicz R. Accuracy of vital status ascertainment using the Social Security Death Master File in a deceased population. *J Insur Med*. 2012;43(3):135-144.
29. Rogers H. H.J.Res.59 - 113th Congress (2013-2014): Continuing Appropriations Resolution, 2014. U.S. Congress; 2013. <https://www.congress.gov/bill/113th-congress/house-joint-resolution/59/text?overview=closed&r=9>. Accessed September 20, 2018.
30. Data Collection and Comparison for ObituaryData.com. <https://www.obituarydata.com/stats.asp>. Accessed September 18, 2018.
31. Centers for Disease Control and Prevention|National Center for Health Statistics. National Death Index. <https://www.cdc.gov/nchs/ndi/index.htm>. Accessed September 18, 2018.
32. CentersforMedicare&MedicaidServices.NewMedicarecards.2018. <https://www.cms.gov/Medicare/New-Medicare-Card/>. Accessed September 18, 2018.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Levin MA, Lin H-M, Prabakar G, McCormick PJ, Egorova NN. Alive or dead: Validity of the Social Security Administration Death Master File after 2011. *Health Serv Res*. 2019;54:24-33. <https://doi.org/10.1111/1475-6773.13069>