

RESEARCH ARTICLE

For reprint orders, please contact: reprints@futuremedicine.com

The value of linking hospital discharge and mortality data for comparative effectiveness research

Background: Linkage of US state hospital discharge records to state death certificate records offers the possibility of tracking long-term mortality outcomes across large, diverse patient populations, which may be useful for comparative effective analyses. **Aim:** To demonstrate the value of linking state community hospital discharge data to vital statistics death files for research by conducting a comparative effectiveness analysis. **Methods:** Linked Patient Discharge Data and Vital Statistics Death Files from the California Office of Statewide Health Planning and Development were used to compare survival rates for patients with an elective repair for abdominal aortic aneurysm who received open aneurysm repair (OAR) versus endovascular aneurysm repair (EVAR). The sample consisted of 13,652 hospitalized patients who underwent an OAR or EVAR for abdominal aortic aneurysm between 1 July 2000 and 31 January 2006. Patients were matched using propensity scores (8966 patients in the matched sample). In-hospital, 30-day, 1-year and 5-year mortality rates were compared between the OAR and EVAR populations, before and after propensity score matching. **Results:** We found a few data anomalies (92 out of 13,652), primarily in patients' sex and date of death. The analysis revealed that in the matched cohort, in-hospital and 30-day postdischarge mortality rates were significantly lower following EVAR than OAR; however, consistent with previous clinical trials, differences in the 1- and 5-year rates were not statistically significant. **Conclusion:** The study demonstrates that linked US state discharge and mortality data can be a valuable resource for comparative effectiveness analyses. In particular, this approach may be useful when generally available data sets such as Medicare claims data limit the generalizability of findings. Policy-makers and others should consider greater investments in these data.

KEYWORDS: abdominal aortic aneurysm ■ comparative effectiveness ■ linked data ■ observational data analysis

Record linkage is widely used in health services, and epidemiologic and outcomes research to obtain comprehensive and robust information on patients, practices and outcomes [1]. In particular, vital statistics data on mortality information have been linked to a large variety of data sets, such as Medicare claims and the National Nursing Home Survey. In the USA, information on mortality is collected by all 50 states and the District of Columbia [2]. The data are standardized by jurisdiction and contributed to the nation's Vital Statistics Cooperative Program under the auspices of the CDC. However, to date, vital statistics mortality data have been linked to hospital discharge records in a limited number of states with limited availability, or only for particular payers (e.g., Medicare) or populations (e.g., registries of patients with particular diseases). This represents a missed opportunity for comparing, analyzing or tracking long-term outcomes of clinical interventions.

For example, Medicare patients comprise approximately 66% of discharges from US community hospitals annually (i.e., ~26 out of 39 million discharges;



Journal of Comparative Effectiveness Research

Tami L Mark*¹, William Lawrence²,
Rosanna M Coffey¹, Timothy Kenney³,
Bong Chul Chu¹, Emile R Mohler 3rd^{4,5}
& Claudia Steiner²

¹Truven Health Analytics, 4301 Connecticut Avenue, NW, Suite 330, Washington, DC 20008, USA

²Agency for Healthcare Research & Quality, 540 Gaither Road Rockville, MD 20850, USA

³Kenney IS Consulting, Inc., 109 North Montgomery Street, Suite D, Ojai, CA 93023, USA

⁴Translational Research Center, Mail Stop 5159, 3400 Civic Center Blvd, Bldg 421, Room 11-103, Philadelphia, PA 19104-15159, USA

⁵Perelman School of Medicine at University of Pennsylvania, PA, USA

*Author for correspondence:

Tel.: +1 301 547 4398

Fax: +1 301 547 4690

tami.mark@truvenhealth.com

author's calculations [101]). Thus, comprehensive and systematic information on postdischarge mortality is absent for approximately one-third of all hospital discharges in the USA. This is a critical omission because the incidence of disease differs by age and the appropriateness and effectiveness of interventions may also differ by age. A comprehensive evaluation of hospital outcomes may be biased if patients without Medicare are omitted.

State hospital discharge databases, linked to state vital statistics, offer a relatively ready source to fill the gap in comprehensive information on postdischarge mortality for all hospital discharges. A recent paper by Bradley and colleagues highlights that state systems are among the high-priority areas that should be leveraged to enhance US healthcare data [3]. California is among the few states that have linked their discharge database – the California Office of Statewide Health Planning and Development Patient Discharge Database – to their vital statistics mortality data – the California State Death Statistical Master File [4]. The linkage procedures used to create this database, and its general strengths and limitations are described by Zingmond and colleagues in a 2004 publication [5]. The authors conclude that the "...identified linkages appear to be highly accurate, and the number of unidentified record linkages has decreased in more recent years of data..."

Although the type of data analysis conducted by Zingmond *et al.* is critical, another way to demonstrate the value of the data is to employ them in an empirical study. By using the data, researchers may further demonstrate to stakeholders, such as state and federal government agencies, the potential value and limitations of the data. This was the motivation behind the present analysis, which examined the ability of the data to be used for comparative effectiveness research. As Meyer and colleagues recently noted, the foundation of comparative effectiveness is understanding effectiveness in the context of large, heterogeneous populations, and comparative effectiveness research "stands to benefit immeasurably" from the growing availability of secondary data [6].

In particular, we decided to focus on the comparative effectiveness of open versus endovascular repair of abdominal aortic aneurysm (AAA) using linked discharge and mortality data. AAA is an outward ballooning of the distal

aorta that supplies blood to the abdomen, pelvis and legs. When detected (typically through diagnostic tests), the treatment goal is to perform surgery before complications or symptoms develop, because aneurysms that cause symptoms require emergency surgery to prevent rupture and complications. Surgery is recommended for patients with aneurysms larger than 5.5 cm in diameter and aneurysms that rapidly increase in size. In a traditional open aortic repair (OAR), a large incision is made in the abdomen and the aneurysmal aorta is replaced with a graft made of synthetic material. Endovascular aortic repair (EVAR) is a newer form of AAA surgery that is less invasive. It uses an endovascular stent graft to reinforce the wall of the aorta to help prevent the aneurysm from rupturing [7]. Prior evidence from randomized controlled trials and observational data find lower all-cause inpatient and 30-day postdischarge mortality for EVAR over OAR, but there are no reported differences in the longer-term benefits of either surgical procedure [8–10].

To summarize, the goal of this investigation is to demonstrate the value of linked, all-payer, community hospital discharge data for all adults to vital statistics death files for comparative effectiveness research by conducting an analysis focused on comparing open versus endovascular repair of AAA.

Methods

■ Data source

We used the California Office of Statewide Health Planning and Development Patient Discharge Database, which consists of records for all discharges from all nonfederal hospitals located in California. The records are linked to the California State Death Statistical Master File, a database of death certificates for all California residents dying in or out of the state. For any given year, approximately 3.5 million discharges are reported within the Patient Discharge Database, and there are 240,000 reported deaths within the Death Statistical Master File.

The hospital discharge and death files were linked by the California Office of Statewide Health Planning and Development using an automated linkage algorithm, primarily created with the patient's Social Security number (SSN) [4]. This linked data file is called 'Linked Patient Data and Vital Statistics Death Statistical Master File'. According to Zingmond *et al.*, approximately 20–30% of patient discharge records do not have

a valid SSN and are not matched [5]. The majority of the records lacking SSNs (~65%) were for births, and the availability of SSN improved in later years. In addition, residents dying outside of California did not have SSNs and were not linked.

We obtained linked Patient Discharge Data and Vital Statistics Death Files from OSPHD from 2000 to 2007 for any individual with an International Classification of Diseases Ninth Revision Clinical Modification (ICD-9-CM) diagnosis code of 441.xx (aortic aneurysm and dissection) appearing anywhere in the diagnosis array. Ultimately, we focused on a subset of these patients with the more specific diagnosis of abdominal aneurysm without mention of rupture as a principal diagnosis (ICD-9-CM code, 441.4). We focused on this group of patients so that the analysis targets patients with planned aneurysm repairs rather than high-risk emergent replacements with a higher risk of in-hospital mortality.

The study design was reviewed and approved by the Committee for the Protection of Human Subjects – the institutional review board for the State of California.

■ Data sample

Out of the 379,334 discharge records in the linked database with ICD-9-CM diagnosis code 441.xx appearing anywhere in the diagnosis array, we identified 24,295 unique patients with abdominal aneurysm without mention of rupture as a principal diagnosis (ICD-9-CM code 441.4).

Table 1 lists the exclusion criteria and their effect on the sample size. We excluded patients with an AAA OAR or EVAR admission in the first 6 months of 2000 to increase the probability of identifying the patient's first OAR rather than a revision of a repair. We also excluded patients with these procedures in 2007 – the last year for which we had data – to ensure that patients could be followed in the data for at least 1 year. These two exclusion criteria reduced the sample by 4754 patients. All patients were required to have ICD-9-CM procedure codes indicating either OAR (38.44 or 39.25) or EVAR (39.71), which reduced the sample by 4058 patients. Patients were excluded if they had both OAR and EVAR procedures during the same admission because we could not attribute the effectiveness to a particular intervention (114 patients met this exclusion criteria). In order to improve the

homogeneity of the two cohorts, we also excluded patients with procedural codes for repair of the thoracic aorta (38.35, 38.45 or 39.73) and/or visceral or renal bypass (38.46, 39.24 or 39.26), polyarteritis nodosa (446.0), coarctation of the aorta (747.1) or Marfan syndrome and other congenital anomalies (759.8). These criteria resulted in a total of 744 excluded patients. A total of 92 patients were excluded owing to anomalies in the data set, including patients who had multiple death records, a death date prior to an admission date, or multiple values for sex. Finally, we excluded 881 patients with OAR or EVAR procedures who were admitted through the emergency room, because they are more likely to have AAA rupture and the study focus was on elective AAA repair. All exclusions were implemented to identify a cohort of reasonably homogenous patients with a similar risk of mortality following these procedures. The trade-off in implementing these exclusions to improve homogeneity and interval validity is that it may reduce the study's generalizability. In particular, the results may not be applicable to patients with AAA rupture or with the excluded comorbid conditions.

The final patient count for this study was 13,652; of these, 7606 had an OAR procedure and 6046 had an EVAR procedure.

■ Measures

The primary outcome measure was mortality, which was measured during the inpatient stay and at 30 days, 1 year and 5 years following discharge, and was obtained from the vital statistics records. The other variables that were used in the logistic regression were obtained from the hospital discharge records. The broad categories of patient characteristics were: sociodemographic characteristics (sex, age, race, ethnicity, median household income for the patient's ZIP code and primary health insurance type); year of first OAR or EVAR AAA procedure; whether or not the patient had a hospitalization with a principal AAA diagnosis in the prior 6 months; whether or not the patient had a hospitalization for any other cause in the prior 6 months; and comorbidities. Two sets of criteria were used to identify comorbid conditions associated with AAA: one suggested by Schwarze *et al.* [11] and the other calculated using the Healthcare Cost and Utilization Project Comorbidity Software [102]. We used the version of this software

Table 1. Sample selection criteria and resulting sample size.

Patient status	Patient count	Cumulative patient count
All discharges in 2000–2007 with the ICD-9 code of '441.4' in the principal diagnosis position	24,295	24,295
Admission date in the first 6 months or the last year of study period covered for OAR and EVAR by the data	4754	19,541
Discharge did not have a code for either OAR or EVAR	4058	15,483
Discharge had an OAR and an EVAR on the same admission	114	15,369
Discharge had 'repair of thoracic aorta' and/or 'visceral renal bypass' procedure code. Discharge had a diagnosis for 'coarctation', 'Marfan syndrome and other congenital anomalies' or 'polyarteritis'	744	14,625
Patient had multiple death records; patient death date is prior to admission date; patient had multiple values of sex	92	14,533
Patient was admitted through the hospital emergency department	881	13,652
Total		
Patient count for OAR treatment group		7606
Patient count for EVAR treatment group		6046

EVAR: Endovascular aneurysm repair; ICD-9: International Classification of Diseases Ninth Revision; OAR: Open aneurysm repair.

that was appropriate to the set of Medicare Severity-Diagnosis Related Group codes in each database. Hospital characteristics, specifically bed size and teaching status, were based on the American Hospital Association Annual Survey data [12]. We calculated from the discharge data the volume of AAA OAR and EVAR repairs conducted over the study period at the hospital where the patient was treated.

■ Analysis

We compared the characteristics of the OAR and EVAR cohorts using means and frequencies, and determined the statistical significance of group differences using χ^2 and t-statistics. The two cohorts were matched using propensity scores for the purpose of obtaining homogenous samples with which to compare outcomes. To derive the propensity score, a logistic model calculated the conditional probability of being in the OAR versus the EVAR cohort.

The predicted propensity score was then used to create a 1:1 match between the OAR and EVAR cohorts using greedy matching with a caliper of 0.25 standard deviation of the propensity score. Balance on the covariates was assessed by computing standardized differences for each covariate; 'balanced' was defined as <10 standardized differences. Finally, we conducted Kaplan–Meier survival analyses on the unmatched and propensity score matched cohorts. t-tests were used to test for statistical significance ($p < 0.05$).

Results

■ Patient & hospital characteristics

Table 2 contains the patient and hospital characteristics for individuals undergoing EVAR or OAR, before and after propensity score matching. After propensity score matching, the cohorts were considered balanced (all standardized differences were <10).

Patient characteristics before matching

Patients undergoing EVAR were more likely than patients undergoing OAR to be male (87 vs 79%), to be older, to be covered by Medicare (83 vs 76%) and to have a higher median household income in their ZIP code area (US\$59,462 vs US\$55,137). Patients having EVAR were less likely to be covered by private health insurance (15 vs 20%). Regarding comorbidities, patients undergoing EVAR were more likely to have diabetes without chronic complications (14 vs 12%) or ischemic heart disease (45 vs 41%), but they were less likely to have chronic pulmonary disease (27 vs 36%), coagulopathy (2 vs 7%), fluid and electrolyte disorders (5 vs 17%), deficiency anemias (7 vs 13%) or preoperative renal insufficiency (6 vs 11%).

Hospital characteristics before matching

Compared with OAR procedures, EVAR procedures were more frequently performed at large hospitals (74 vs 69%), teaching hospitals (52 vs 35%) and hospitals with a higher average number of AAA OAR and EVAR procedures (105 vs 72).

Table 2. Baseline patient and hospital characteristics of individuals undergoing endovascular aortic repair or open aortic repair of abdominal aortic aneurysm, before and after propensity score matching.

Variables	Unmatched cohort		p-value	Matched cohort		Standardized difference	p-value
	EVAR (n = 6046)	OAR (n = 7606)		EVAR (n = 4483)	OAR (n = 4483)		
Sex							
Female (%)	13	21	<0.001	15	16	3.38	0.11
Age group (years, %)							
0–54	1	2	<0.001	1	1	0.59	0.78
55–64	10	15	<0.001	11	13	4.01	0.06
65–74	34	40	<0.001	37	38	2.58	0.22
75–84	45	39	<0.001	43	42	1.76	0.41
≥85	10	5	<0.001	8	6	7.01	<0.001
Primary insurance (%)							
Medicare	83	76	<0.001	81	79	3.34	0.11
Medi-Cal (Medicaid)	1	2	0.008	1	2	1.31	0.54
Private insurance	15	20	<0.001	17	18	2.94	0.17
Other	1	2	0.026	1	1	0.61	0.77
Race (%)							
White	85	88	<0.001	87	87	2.06	0.33
Black	3	3	0.089	3	3	0.28	0.90
Asian/Pacific Islander	5	6	0.305	6	6	0.19	0.93
Other	7	3	<0.001	4	4	3.30	0.12
Hispanic ethnicity	6	6	0.990	6	6	0.29	0.89
Income (US\$)							
Median household income for patient's ZIP code, mean (SD)	59,462 (22,844)	55,137 (20,866)	<0.001	58,011 (21,622)	57,141 (21,829)	4.00	0.06
Year of procedure (%)							
2000	2	10	<0.001	2	2	0.43	0.84
2001	11	21	<0.001	13	15	4.79	0.02
2002	12	18	<0.001	15	18	6.63	0.00
2003	15	14	0.620	16	17	3.18	0.13
2004	18	15	<0.001	19	18	1.33	0.53
2005	20	12	<0.001	18	16	5.45	0.01
2006	22	10	<0.001	17	14	7.74	<0.001
Hospital size (%)							
Large bed size hospital	74	69	<0.001	73	73	0.45	0.83
Teaching hospital status (%)							
Teaching hospital	52	35	<0.001	45	42	7.25	<0.001
Hospital procedures, mean (SD)							
Average number of abdominal aortic repairs at hospital	105 (85)	72 (46)	<0.001	88 (70)	82 (49)	10.58	<0.001

EVAR: Endovascular aortic repair; OAR: Open aortic repair; SD: Standard deviation.

Table 2. Baseline patient and hospital characteristics of individuals undergoing endovascular aortic repair or open aortic repair of abdominal aortic aneurysm, before and after propensity score matching (cont.).

Variables	Unmatched cohort		p-value	Matched cohort		Standardized difference	p-value
	EVAR (n = 6046)	OAR (n = 7606)		EVAR (n = 4483)	OAR (n = 4483)		
Comorbidities (%)							
Congestive heart failure	0	1	<0.001	0	0	2.11	0.32
Other neurological disorders	2	2	0.970	2	2	0.34	0.87
Chronic pulmonary disease	27	36	<0.001	29	31	2.87	0.18
Diabetes without chronic complications	14	12	<0.001	14	13	1.65	0.44
Diabetes with chronic complications	1	1	0.130	1	1	1.14	0.59
Hypothyroidism	6	6	0.935	6	6	0.09	0.97
Renal failure	5	5	0.165	5	5	0.64	0.76
Liver disease	1	2	0.005	1	1	0.21	0.92
Peptic ulcer disease	1	1	0.001	1	1	1.62	0.44
Lymphoma	1	1	0.370	1	1	0.27	0.90
Metastatic cancer	1	0	0.016	1	0	1.28	0.55
Solid tumor without metastasis	7	8	0.019	7	7	1.20	0.57
Coagulopathy	2	7	<0.001	3	3	3.37	0.11
Obesity	7	6	0.626	7	7	0.35	0.87
Weight loss	0	2	<0.001	1	1	1.18	0.58
Fluid and electrolyte disorders	5	17	<0.001	6	7	3.40	0.11
Chronic blood loss anemia	0	1	<0.001	0	1	0.32	0.88
Deficiency anemias	7	13	<0.001	8	8	1.55	0.46
Psychoses	1	1	<0.001	1	1	1.74	0.41
Depression	2	2	0.061	2	2	1.12	0.60
Preoperative renal insufficiency	6	11	<0.001	7	7	1.86	0.38
Ischemic heart disease	45	41	<0.001	43	43	0.09	0.97
Cerebral vascular occlusive disease	5	5	0.239	5	5	0.32	0.88
Hospitalization (%)							
In prior 6 months	23	22	0.105	22	22	0.11	0.96
In prior 6 months with abdominal aortic aneurysm diagnosis	17	16	0.033	17	17	0.60	0.78

EVAR: Endovascular aortic repair; OAR: Open aortic repair; SD: Standard deviation.

Characteristics after propensity score matching

The sample was considered balanced after propensity score matching because almost all of the baseline characteristics had a standardized difference of 10 or less between the OAR and EVAR cohorts. The exceptions were: the average number of AAA repairs completed at the hospital, which was slightly higher for EVAR compared with OAR discharges (88 vs 82); and whether or not the repairs were completed at a teaching hospital (45 vs 42% for EVAR and OAR, respectively).

Mortality outcomes

In the matched sample, in-hospital and 30-day postdischarge mortality rates were significantly lower for patients undergoing EVAR than for patients undergoing OAR (1.2 vs 2.5% and 1.5 vs 2.8%, respectively). For 1- and 5-year mortality in the matched EVAR and OAR cohorts, the difference in the probability of death between cohorts was not statistically significant (Table 3 & Figure 1). The size of the differences in mortality rates between EVAR and OAR was significantly reduced after matching (Figure 1 vs Figure 2).

Table 3. Mortality outcomes of individuals undergoing endovascular aortic repair or open aortic repair of abdominal aortic aneurysm, before and after propensity score matching.

Variable	Unmatched cohort		p-value	Matched cohort		p-value
	EVAR (n = 6046), % (95% CI)	OAR (n = 7606), % (95% CI)		EVAR (n = 4483), % (95% CI)	OAR (n = 4483), % (95% CI)	
In-hospital mortality	0.98 (0.76–1.26)	4.40 (3.97–4.89)	<0.001	1.20 (0.92–1.57)	2.52 (2.10–3.02)	<0.001
30-day postdischarge mortality	1.24 (0.99–1.55)	4.71 (4.25–5.21)	<0.001	1.45 (1.14–1.85)	2.81 (2.37–3.34)	0.035
1-year mortality	2.55 (2.18–2.98)	5.44 (4.96–5.98)	0.021	2.83 (2.39–3.36)	3.52 (3.02–4.11)	0.085
5-year mortality	8.06 (7.21–9.00)	8.76 (8.10–9.46)	0.465	7.99 (7.07–9.01)	6.81 (6.02–7.70)	0.577

EVAR: Endovascular aortic repair; OAR: Open aortic repair.

Discussion

The goal of this study was to conduct a comparative effectiveness analysis with an all-age, all-payer discharge database that had been linked to postdischarge mortality information, to examine the potential value of such data. We draw observations from these efforts regarding the usefulness of linked discharge–mortality databases for comparative effectiveness research and the potential applications of a linked discharge–mortality database for other types of research.

■ Potential usefulness of linked data for comparative effectiveness research

We found the linked database relatively easy to use for observational comparative effectiveness analysis. Importantly, our results are consistent with those of prior randomized clinical trials that find an advantage of EVAR over OAR with respect to immediate inpatient and perioperative mortality, but no longer-term advantage [13,14].

Two studies have also compared EVAR with OAR using Medicare claims linked to vital statistics mortality data. Schermerhorn and colleagues, using data for 2001–2004, reported an advantage of EVAR over OAR in survival up until 3 years, at which time the survival rates converge [15]. By contrast, Jackson and colleagues, using 2003–2007 data, found a survival advantage for EVAR that was maintained throughout the 5.7-year follow-up [16]. Our analysis is consistent with these two prior studies in showing a perioperative survival advantage of EVAR over OAR. The absolute 30-day mortality differences are also similar; the advantage of EVAR over OAR was 1.2 versus 4.7% (present study), 1.2 versus 4.8% [15] and 1.2 versus 5.8% [16].

In our propensity score-matched analyses, the survival difference is not significant after

1 year, whereas in the unmatched analyses the survival advantage lasts beyond 1 year. Lack of a survival advantage after 1 year is consistent with results from randomized clinical trials, but not with the two prior studies using Medicare claims. Additionally, our unmatched results indicate a long-term survival advantage. Such discrepancies in long-term survival advantage of EVAR over OAR suggest that the longer-term outcomes may vary owing to patient heterogeneity (e.g., age and comorbidities) or other factors (e.g., surgical techniques) that need to be more fully understood. For example, the descriptive statistics comparing the EVAR and OAR populations reveal that EVAR tends to be more commonly used among older patients who may be less able to tolerate an open surgery. In general, these findings support the assertion that observational data are a cost-effective supplement

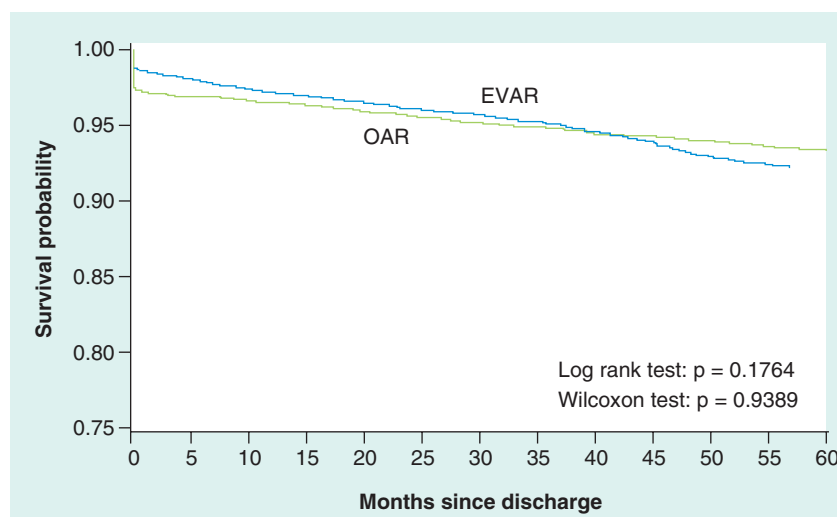


Figure 1. Kaplan–Meier plot for the 5-year survival by endovascular aortic repair or open aortic repair of abdominal aortic aneurysm (propensity score-matched cohort).

EVAR: Endovascular aortic repair; OAR: Open aortic repair.

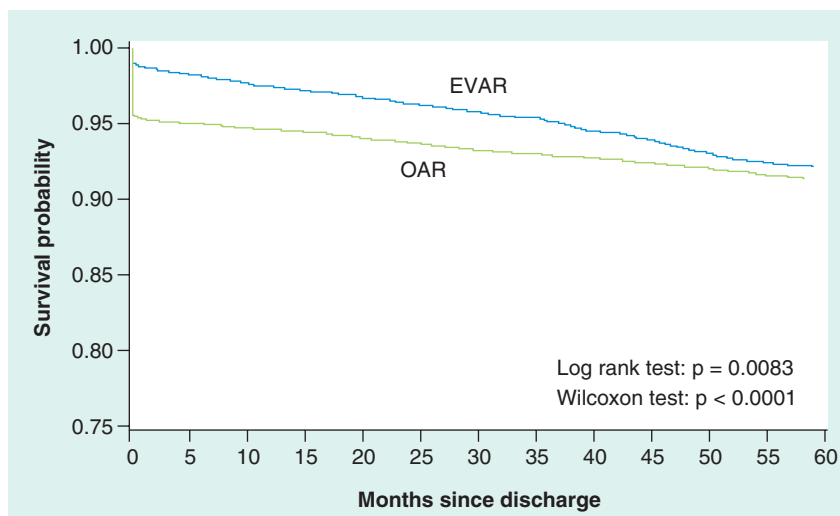


Figure 2. Kaplan–Meier plot for the 5-year survival by endovascular aortic repair or open aortic repair of abdominal aortic aneurysm (unmatched cohort).

EVAR: Endovascular aortic repair; OAR: Open aortic repair.

to clinical trials to evaluate interventions in real-world settings and a potential alternative when randomized clinical trials are infeasible.

Clearly, the data will only be useful for some types of comparative effectiveness analyses. In particular, it will be most useful when evaluating interventions that are provided primarily in inpatient hospital settings for which ICD-9-CM procedure codes or other standardized codes exist to identify the interventions in the discharge records. Additionally, the data will be useful for comparative effectiveness studies where the conditions, procedures and posthospital mortality being evaluated are sufficiently prevalent to provide adequate statistical power to detect differences in mortality outcomes. In contrast to Medicare claims data, which primarily capture data for individuals aged 65 years and older, discharge data capture all payers. Given that approximately 66% of individuals discharged from community hospitals are younger than 65 years, linked discharge data can be particularly informative for researching interventions relevant to this population. These include percutaneous transluminal coronary angioplasty, spinal fusion, laminectomy, hip replacement and prostatectomy procedures (authors' calculation of frequent procedures among populations younger than 65 years using [101]).

As with all observational data, the linked discharge–mortality data will face the issue of unobserved confounding from selection bias. The ability of researchers to address confounding is

somewhat more limited with linked discharge and mortality data – as compared with claims data linked to mortality data – because the linked discharge–mortality data lack information on prescription drug and outpatient utilization that can be used for statistical adjustment. Lack of outpatient data also precludes examination of outpatient interventions, either as alternatives to hospital interventions or as indications of complicated cases. A second limitation of the linked discharge–mortality data is the lack of clinical detail. For example, for this particular study of AAA procedures, information about aneurysm size or anatomical features was not available in the hospital discharge records. However, the Medicare claims data include the same limitation, and they have continued to be utilized for observational studies. The third common criticism of administrative data concerns errors in coding of diagnoses or procedures. We conducted a literature search and did not find any published studies that examined the validity of ICD-9 coding of AAA or OAR and EVAR, although we have no reason to believe that the data were incorrectly coded.

We discovered a small number of data errors from the OSPHD database across the 5 years; there were 92 erroneous records out of a total of 13,652 (0.67%). These errors primarily represented individuals who were listed as male on one discharge record and female on another. In a few cases, there were multiple death records, and reports of deaths that occurred prior to an admission. Dr Zigmond and OSPHD staff members who conducted the data linkage reported that errors may occur either because miscoded SSNs slip through the linkage and validation process or because of fraudulent uses of another individual's SSN upon admission to a hospital. These small errors did not appear to have a significant impact on our ability to conduct the study, but this issue needs to be evaluated further. Another consideration is that the linkage process excludes patients without SSNs. The majority of missing records are for young children and new births; however, the majority of missing deaths are for older adults. More research is needed to understand the implications of these missing records for introducing biases or limiting the generalizability of the results. Moreover, if government agencies start to place more limits on SSNs for vital statistics, the deterministic approach used to link hospital discharge and mortality data may not be viable, and alternative

methods such as probabilistic linking may need to be employed [103].

■ Potential usefulness of linked data for other research

In addition to observational comparative effectiveness studies, linked discharge–mortality data may be of value for a number of other purposes. First, the data can be used to understand the epidemiology of diseases or treatments, such as incidence rates of particular outcomes (e.g., rates of conversion from EVAR to OAR, rates of complication and postdischarge mortality rates) overall or for different subpopulations of patients. This information may be helpful in: anticipating and mitigating clinical complications of the procedures for certain subgroups; discharge planning and timing of follow-up monitoring for adverse outcomes; avoiding a specific type of procedure in some complicated cases; developing hypotheses and planning further research on the condition and alternative treatments using more rigorous research designs; and examining analyses within subgroups, in particular for individuals not covered by Medicare.

Second, another potential application is in pragmatic randomized clinical trials where data from administrative records linked to vital statistics are used to track outcomes following randomization. Examples of such ‘naturalistic’ and ‘streamlined’ studies are available from the UK, The Netherlands and Denmark, among other countries [104]. For example, in the FAST trial being conducted in the UK and Denmark, medications are provided to patients by mail, and hospitalization and mortality outcomes are monitored via linked records. The usefulness of the data for this type of research is predicated on the feasibility of establishing processes to identify patient records.

A third usage may be to improve knowledge of hospital-level outcomes and practice patterns. For

example, the state hospital discharge data capture all procedures conducted at the hospital; thus, one can create measures of total hospital volume and compare them with postdischarge mortality, which has only been feasible to date for the subgroup of Medicare beneficiaries. For example, OSPHD has used the data to compare heart attack mortality rates for Californian hospitals [17,18].

Conclusion

In summary, the present study demonstrates that linking state discharge and vital statistics mortality records offers a promising opportunity for comparative effectiveness research, as well as other research areas. Investments in linking state discharge and vital statistics mortality records can be a cost-effective stimulus to observational research as a companion to clinical trials. Population-based observational studies using statewide discharge data linked to vital statistics mortality data can serve as an early indicator of benefit or harm that can generate robust clinical trials. Additionally, the data can assess real-world effectiveness once comparative effectiveness clinical trials are introduced to the general population.

Disclaimer

The authors of the report are responsible for its content. Statements in the report should not be construed as endorsement by the Agency for Healthcare Research and Quality or the US Department of Health and Human Services.

Financial & competing interests disclosure

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.

Executive summary

- Linkage of US state hospital discharge records to state death certificate records offers the possibility of tracking long-term mortality outcomes across large, diverse patient populations, which may be useful for comparative effective analyses.
- The value of linked state community hospital discharge and vital statistics death files was demonstrated by using the Patient Discharge Data and Vital Statistics Death Files from the California Office of Statewide Health Planning and Development to demonstrate in a comparative effectiveness analysis of survival rates of patients with an elective repair for abdominal aortic aneurysm who received open aneurysm repair versus endovascular aneurysm repair.
- The study demonstrated that linked US state discharge and mortality data can be a valuable resource for comparative effectiveness analyses. In particular, this approach may be useful when generally available data sets, such as Medicare claims data, limit the generalizability of findings.
- Policy-makers and others should consider greater investments in these data.

References

- 1 Roundtable on evidence-based medicine. summary. In: *The Learning Healthcare System: Workshop Summary (IOM Roundtable on Evidence-Based Medicine)*. Olsen L, Aisner D, McGinnis JM (Eds). The National Academies Press, Washington, DC, USA, 1–36 (2007).
- 2 Siri MJ, Cork DL. *Vital Statistics: Summary of a Workshop*. Committee on: National Statistics, Division of Behavioral and Social Sciences and Education, National Research Council of the National Academies. The National Academies Press, Washington, DC, USA (2009).
- 3 Bradley CJ, Penberthy L, Devers KJ, Holden DJ. Health services research and data linkages: issues, methods, and directions for the future. *Health Serv. Res.* 45(5 Pt 2), 1468–1488 (2010).
- 4 Office of Statewide Health Planning and Development, Healthcare Information Division. *Linked Patient Discharge Data and Vital Statistics Death Statistical Master File 2000–2007*. Healthcare Information Division of the State of California, CA, USA.
- 5 Zingmond DS, Ye Z, Ettner SL, Liu H. Linking hospital discharge and death records – accuracy and sources of bias. *J. Clin. Epidemiol.* 57(1), 21–29 (2004).
- 6 Meyer AM, Carpenter WR, Abernethy AP, Stürmer T, Kosorok MR. Data for cancer comparative effectiveness research: past, present, and future potential. *Cancer* 118(21), 5186–5197 (2012).
- 7 Hirsch AT, Haskal ZJ, Hertzner NR *et al.* ACC/AHA Guidelines for the Management of Patients with Peripheral Arterial Disease (lower extremity, renal, mesenteric, and abdominal aortic): a collaborative report from the American Associations for Vascular Surgery/Society for Vascular Surgery, Society for Cardiovascular Angiography and Interventions, Society for Vascular Medicine and Biology, Society of Interventional Radiology, and the ACC/AHA Task Force on Practice Guidelines (writing committee to develop guidelines for the management of patients with peripheral arterial disease) – summary of recommendations. *J. Vasc. Interv. Radiol.* 17(9), 1383–1397 (2006).
- 8 Greenhalgh RM, Brown LC, Kwong GP *et al.* Comparison of endovascular aneurysm repair with open repair in patients with abdominal aortic aneurysm (EVAR trials 1), 30-day operative mortality results: randomized controlled trial. *Lancet* 364, 843–848 (2004).
- 9 Prinssen M, Buskens E, de Jong SE *et al.* DREAM trial participants. Cost–effectiveness of conventional and endovascular repair of abdominal aortic aneurysms: results of a randomized trial. *J. Vasc. Surg.* 46(5), 883–890 (2007).
- 10 EVAR trial participants. Endovascular aneurysm repair versus open repair in patients with abdominal aortic aneurysm (EVAR trial 1): randomised controlled trial. *Lancet* 365(9478), 2179–2186 (2005).
- 11 Schwarze ML, Shen Y, Hemmerich J *et al.* Age-related trends in utilization and outcome of open and endovascular repair for abdominal aortic aneurysm in the United States, 2001–2006. *J. Vasc. Surg.* 50(4), 722–729 (2009).
- 12 *American Hospital Association Annual Survey*. Health Forum, LLC, American Hospital Association, Chicago, IL, USA (2012).
- 13 Lederle FA, Freischlag JA, Kyriakides TC *et al.* Outcomes following endovascular vs. open repair of abdominal aortic aneurysm: a randomized trial. *JAMA* 302(14), 1535–1542 (2009).
- 14 Blankensteijn JD, de Jong SE, Prinssen M *et al.* Two-year outcomes after conventional or endovascular repair of abdominal aortic aneurysms. *N. Engl. J. Med.* 352, 2398–2405 (2005).
- 15 Schermerhorn ML, O’Malley AJ, Jhaveri A *et al.* Endovascular vs. open repair of abdominal aortic aneurysms in the Medicare population. *N. Engl. J. Med.* 358(5), 464–474 (2008).
- 16 Jackson RS, Chang DC, Freischlag JA. Comparison of long-term survival after open vs endovascular repair of intact abdominal aortic aneurysm among Medicare beneficiaries. *JAMA* 307(15), 1621–1628 (2012).
- 17 Zach AP, Romano PS, Luft HS. *Report on Heart Attack 1991–1993, Volume 1: User’s Guide*. California Office of Statewide Health Planning and Development, Sacramento, CA, USA (1997).
- 18 Healthcare Quality and Analysis Division. *Report on Heart Attack Outcomes in California 1996–1998, Volume 2: Technical Guide*. California Office of Statewide Health Planning and Development, Sacramento, CA, USA (2002).

■ Websites

- 101 US Department of Health & Human Services. Welcome to HCUPnet. <http://hcupnet.ahrq.gov>
- 102 HCUP Comorbidity Software. Healthcare Cost and Utilization Project (HCUP). Agency for Healthcare Research and Quality, Rockville, MD, USA (2012). www.hcup-us.ahrq.gov/toolssoftware/comorbidity/comorbidity.jsp
- 103 Sack K. Researchers wring hands as US clamps down on death record access. www.nytimes.com/2012/10/09/us/social-security-death-record-limits-hinder-researchers.html?_r=0
- 104 MacDonald T, Mackenzie I, Wei L. Novel ways to get good trial data: the UK experience. Medical Research Institute of Dundee. Institute of Medicine of the National Academies. www.iom.edu/Home/Global/Perspectives/2012/GoodTrialDataUK.aspx