



Published in final edited form as:

Am Heart J. 2009 June ; 157(6): 995–1000. doi:10.1016/j.ahj.2009.04.002.

Linking Inpatient Clinical Registry Data to Medicare Claims Data Using Indirect Identifiers

Bradley G. Hammill, MS^a, Adrian F. Hernandez, MD, MHS^{a,b}, Eric D. Peterson, MD, MPH^{a,b}, Gregg C. Fonarow, MD^c, Kevin A. Schulman, MD^{a,b}, and Lesley H. Curtis, PhD^{a,b}

^a Duke Clinical Research Institute, Duke University School of Medicine, Durham, North Carolina

^b Department of Medicine, Duke University School of Medicine, Durham, North Carolina

^c University of California, Los Angeles Medical Center

Abstract

Background: Inpatient clinical registries generally have limited ability to provide a longitudinal perspective on care beyond the acute episode. We present a method to link hospitalization records from registries with Medicare inpatient claims data, without using direct identifiers, to create a unique data source that pairs rich clinical data with long-term outcome data.

Methods and Results: The method takes advantage of the hospital clustering observed in each database by demonstrating that different combinations of indirect identifiers within hospitals yield a large proportion of unique patient records. This high level of uniqueness also allows linking without advance knowledge of the Medicare provider number of each registry hospital. We applied this method to 2 inpatient databases and were able to identify 81% of 39,178 records in a large clinical registry of patients with heart failure and 91% of 6581 heart failure records from a hospital inpatient database. The quality of the link is high and reasons for incomplete linkage are explored. Finally, we discuss the unique opportunities afforded by combining claims and clinical data for specific analyses.

Conclusions: In the absence of direct identifiers, it is possible to create a high-quality link between inpatient clinical registry data and Medicare claims data. The method will allow researchers to use existing data to create a linked claims-clinical database that capitalizes on the strengths of both types of data sources.

Introduction

In the past decade, several multicenter clinical registries have formed to support quality assessment, quality improvement, and clinical research. These disease- and procedure-specific registries collect data on large numbers of patients throughout the United States and provide a wealth of information about patient characteristics, patterns of inpatient care, and early outcomes.¹⁻³ However, registries are generally limited to in-hospital or 30-day outcomes and thus have limited ability to provide a longitudinal perspective on care beyond the acute episode. Although extension of data collection to outpatient follow-up is desirable, it has been beyond

© 2009 Mosby, Inc. All rights reserved.

Address correspondence to: Lesley H. Curtis, PhD, Center for Clinical and Genetic Economics, Duke Clinical Research Institute, PO Box 17969, Durham, NC 27715; telephone: 919-668-8101; fax: 919-668-7124; e-mail: lesley.curtis@duke.edu..

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

the scope of most clinical registries and requires informed consent that would limit enrollment to smaller cohorts.⁴

Administrative data sets, such as Medicare claims data, are a ready-made alternative for longitudinal outcome assessment. Administrative data that contain patient identifiers allow for the creation of a continuous record of hospitalizations, outpatient care, and outcomes, but they lack detailed clinical information about patients' disease states or treatments. Linking hospitalization records from inpatient clinical registries with Medicare claims data can produce a unique data source that contains rich clinical data paired with long-term outcome and hospitalization data that can be used to inform best practices, provide a surveillance network to evaluate the safety of therapeutics, and stimulate clinical research.

The primary challenge in linking clinical and claims data is that most national registries do not collect or distribute direct patient identifiers. In this paper, we describe a method for identifying records from inpatient clinical registries in Medicare inpatient claims data using indirect identifiers. Indirect identifiers are nonunique fields, such as admission date, discharge date, and patient age or date of birth. The method takes advantage of the hospital clustering observed in each database by demonstrating that different combinations of indirect identifiers within hospitals yield a large proportion of unique patient records. This high level of uniqueness allows linking without advance knowledge of the Medicare provider number of each registry hospital. After describing the linking method, we demonstrate its use in 2 inpatient databases and examine the characteristics of the results.

Methods and Results

Data Sources and Patient Populations

We used 2 inpatient databases in this study, neither of which made direct patient identifiers available to us. First, we used data from the Organized Program to Initiate Life-Saving Treatment in Hospitalized Patients With Heart Failure (OPTIMIZE-HF) registry.⁵ This registry contains information on eligible hospitalizations from hospitals that participated voluntarily in the OPTIMIZE-HF quality improvement program. Eligible hospitalizations included those for which heart failure was the primary cause of admission or those in which significant heart failure symptoms developed during the inpatient stay. For this analysis, we included the 39,178 hospitalizations of patients aged 65 years or older who were discharged between January 1, 2003, and December 31, 2004.

We also used information from the Duke University all-payer inpatient database. This database contains information on all hospitalizations at Duke University Hospital, a 900-bed tertiary and quaternary care teaching hospital in Durham, North Carolina. To mirror the OPTIMIZE-HF population, we extracted data on all hospitalizations for which a diagnosis of heart failure was recorded (International Classification of Diseases, Ninth Revision, Clinical Modification [ICD-9-CM] diagnosis code 428.x, 402.x1, 404.x1, or 404.x3). We retained service dates, ICD-9-CM diagnosis codes, patient demographic characteristics, and payer data and further limited the data set to include only the 6581 hospitalizations of patients aged 65 years or older who were discharged between January 1, 2004, and December 31, 2006.

The administrative claims data source for this study was the 100% Medicare inpatient claims file, which contains information on all hospitalizations of patients enrolled in fee-for-service Medicare and includes service dates and ICD-9-CM diagnosis codes. The database contains anonymous patient identifiers, which enables follow-up of beneficiaries over time but does not enable identification of any beneficiary through their Medicare health insurance number. In addition, the 100% Medicare denominator file, which links to the inpatient file, contains information on beneficiary eligibility, demographic characteristics, and date of death. We used

data on all hospitalizations of beneficiaries aged 65 years or older who were discharged between January 1, 2003, and December 31, 2006. To mirror the OPTIMIZE-HF and Duke populations, we also created a subset limited to hospitalizations for which heart failure was a listed diagnosis (ICD-9-CM diagnosis code 428.x, 402.x1, 404.x1, or 404.x3). There were more than 11 million inpatient claims each year for hospitalizations of elderly patients in the Medicare database, of which approximately 2.5 million had an associated heart failure diagnosis.

The institutional review board of the Duke University Health System approved the study. The Centers for Medicare & Medicaid Services approved the use of the Medicare claims data for the study. The work was supported by grant U18HS10548 from the Agency for Healthcare Research and Quality and a research agreement between GlaxoSmithKline and Duke University. The authors are solely responsible for the design and conduct of this study, all study analyses, the drafting and editing of the paper, and its final contents.

Uniqueness of Records

We began by searching each database for combinations of fields that were unique (or nearly unique), routinely collected, and objectively coded. Available fields for linking included admission date, discharge date, patient sex, and patient date of birth or age. We also employed the hospital identifier as a linking field. In the claims data, we considered the Medicare provider identifier to be the hospital identifier. In the registry data, we considered the site identifier to be the hospital identifier.

We calculated the percentage of records in each data source that were unique given specific combinations of distinct values of the variables. We also checked for unique records using combinations that allowed for flexibility in certain fields. For rules using date of birth, we checked for unique records using combinations that involved any 2 of the 3 components of date of birth (ie, month, day, and year). For rules using age, we allowed ages to differ by 1 year and service dates to differ by 1 day. Only within-hospital results are reported, because no combinations of these variables resulted in more than 85% uniqueness across all records in these databases. Because of the large size of the Medicare claims database, we report the proportion of unique records for the 2004 Medicare data only.

Table 1 shows the within-hospital uniqueness among all claims and in the heart failure subset of claims. Combinations 1 through 12 used date of birth; combinations 13 through 24 used age. Among all claims, almost all records were unique when considering distinct dates of birth along with any combination of admission date and discharge date, regardless of patient sex (combinations 1 through 6). Even when only 2 of the 3 components of date of birth were considered along with both service dates, over 98% of the records were unique (combinations 7 through 8). This high proportion of unique records was true regardless of which 2 components of date of birth were considered (data not shown). When age was considered, combinations of distinct age, sex, and both service dates resulted in 97% uniqueness (combination 13). No other combination involving age resulted in more than 94% uniqueness (combinations 14 through 24).

In the subset of heart failure claims, the percentage of unique records associated with all variable combinations increased. All combinations that included either distinct or partial dates of birth resulted in at least 97% uniqueness (combinations 1 through 12). Combinations of distinct age and both service dates, regardless of patient sex, resulted in more than 98% uniqueness (combinations 13 through 14). Among the rules involving age, even allowing flexibility in age, admission date, or discharge date did not substantially reduce the proportion of records that were unique, as long as sex was considered (combinations 15 through 17). The uniqueness of records in the OPTIMIZE-HF and Duke heart failure databases (not shown) was similar to the results found in the Medicare subset.

Hospital-Level Linkage

Because substantially higher proportions of records were unique within hospitals as compared with across all hospitals, we required that hospitals match between data sources when linking. Registries rarely collect Medicare provider numbers, so we established a hospital “crosswalk” by matching hospitalizations from each inpatient database with Medicare hospitalizations on the basis of exact values of admission date, discharge date, patient sex, and patient date of birth. The Medicare hospital(s) that contained the preponderance of exactly matched records for a given registry hospital were presumed to be the correct link for that registry hospital. We validated these links by comparing the Medicare hospital names with the registry hospital names and by searching for improper links between Medicare hospitals and Veterans Affairs (VA) hospitals. VA hospitals should not be located since they do not receive payment through the Medicare program.

Of the 255 hospitals participating in the OPTIMIZE-HF registry, 208 (81.6%) had at least 5 exact matches with 1 or more Medicare hospitals, and each of these links was correct, as confirmed by the hospital names in both sources. Although most registry hospitals linked to single Medicare hospitals, 6 (2.4%) linked to multiple Medicare hospitals. These were cases for which either the Medicare provider number for the registry hospital changed over the course of enrollment or data from multiple affiliated hospitals were submitted to the registry under a single identifier. None of the 11 VA hospitals participating in OPTIMIZE-HF were identified in the Medicare data. The remaining 36 unidentified registry hospitals enrolled a total of 150 patients, less than 0.5% of the total registry enrollment. Using hospital names and addresses, it was possible to identify Medicare hospital identifiers manually for almost all of these remaining smaller non-VA registry hospitals.

Hospitalization-Level Linkage

Rules for linking between databases were based on the combinations of variables that resulted in > 98% uniqueness in our data. High-volume hospitals tended to have substantially more duplicate records when we considered a lower threshold. Records in each database must match, either exactly or to the given level of flexibility, on the fields indicated by a given rule to be considered a valid link. We linked each of the inpatient databases to the Medicare heart failure claims subset, which yielded consistently higher proportions of unique records. Some registries make age, but not date of birth, available to researchers, so we considered the candidate linking rules in 2 sets, based on whether the rules included date of birth or age. Within each set, we successively applied each rule, from most specific to least specific. For each rule, we calculated how many records from each inpatient database were located in the Medicare data. After applying all rules, we calculated the cumulative number and percentage of records located. Only hospitalizations that did not match for a given rule were passed along to the next rule for matching. We allowed each Medicare hospitalization to link to the single registry record with the best evidence for a match, and registry hospitalizations that could be linked to multiple Medicare records using the same rule were not linked to any Medicare records.

Table 2 shows the results of hospitalization-level linking. Rules 1 through 10 used date of birth; rules 13 through 17 used age. In all, approximately 80% of the OPTIMIZE-HF records and 90% of the records in the Duke database were located in the Medicare data. Using date of birth generally resulted in a slightly higher proportion of matches for each database. For each set of results, a small number of rules identified the large majority of patients in each database.

Evaluation of the Results

We evaluated the links in several ways. Medicare claims data are limited to fee-for-service claims, so linked hospitalizations for patients enrolled in Medicare managed care programs are likely erroneous. In the Duke database, we used payer status to assess the extent to which

Medicare fee-for-service hospitalizations had been linked to managed care hospitalizations. Among the 109 Duke heart failure hospitalizations of patients enrolled in Medicare managed care plans, none were incorrectly linked to the claims data using date of birth and only 1 (0.9%) was incorrectly linked using age.

In an effort to understand why specific hospitalizations of patients enrolled in the Medicare fee-for-service program were not located as expected, we compared link rates for a few specific subgroups from the Duke data. First, it is possible that the claim never reached the Medicare program because it was paid entirely by an employer-based or private payer. Of the 6472 hospitalizations of patients enrolled in fee-for-service Medicare, 309 (4.8%) listed Medicare as the secondary payer. Hospitalizations listing Medicare as the primary payer were located 96% of the time, whereas those listing Medicare as the secondary payer were located only 37% of the time. Second, we examined hospitalizations that listed an eligible ICD-9-CM diagnosis code for heart failure beyond the tenth position in the inpatient database, because only 10 diagnoses are accepted on a Medicare inpatient claim. Only 56 (0.9%) of the 6472 records listed a diagnosis code for heart failure beyond the tenth position in the hospital database. Among those records, 1 (1.8%) linked to Medicare records.

Discussion

Although previous work has been done to link different databases to Medicare claims data,⁶⁻⁸ these efforts required direct identifiers like patient name or Social Security number. In this paper, we describe and demonstrate a method that enables researchers to identify a high proportion of clinical registry database hospitalizations in 100% Medicare inpatient claims data without direct patient identifiers. We found that using combinations of nonunique fields commonly available in both registry and claims databases was sufficient for merging the databases. In the parlance of other linking methods,⁹ we found that blocking by hospital was a critical aspect of any inpatient record linkage, and we demonstrated that it was not necessary to have an existing crosswalk between registry and Medicare hospital identifiers because such a crosswalk could be developed reliably using the data at hand. The Box summarizes the recommended steps needed to link an inpatient registry with Medicare inpatient claims data.

The process of linking the databases involves several important decisions. First, after limiting both databases to include only elderly patients, it is helpful to limit the Medicare data to hospitalizations similar to those included in the registry. A broad list of inclusion criteria is preferable to avoid inadvertently excluding patients who might appear in the registry. It may not be possible to create a subset of the Medicare data when a registry's enrollment is based on a diagnosis that is known to be undercoded in claims data or when the inclusion criteria rely on clinical measures not recorded in the claims data. In such cases, Table 1 suggests that a link between the registry and the entire set of Medicare claims can still be made if the registry includes date of birth.

Second, a crosswalk between hospital identifiers in the registry and the Medicare claims data must be developed. If such a crosswalk already exists, this step may still be useful both to confirm the existing crosswalk and to find additional Medicare identifiers for registry hospitals. As we found in the OPTIMIZE-HF registry, several hospitals changed Medicare provider identifiers during the registry enrollment period and some registry sites enrolled patients from multiple affiliated hospitals. A strength of the method we describe is that definitive links between hospitals in each data source need not be known ahead of time.

Third, the decision must be made whether to allow Medicare records to link to multiple registry records. In this study, we allowed each Medicare record to link only with the registry record with which it had the highest evidence of a match. This approach made sense for both

hospitalization-based databases used in this study because we expected a one-to-one relationship between registry records and Medicare records. If a registry's unit of analysis is a procedure, multiple procedures in a single hospitalization may be recorded separately in the registry. In this case, allowing Medicare records to match multiple registry records may be preferable.

Fourth, linking rules must be selected. If date of birth is present in the registry data, use some or all of rules 1 through 10. If only age is present, use some or all of rules 13 through 17. The results presented in Table 2 indicate that using as few as 3 rules may capture the large majority of links between databases and that there is no need to rely solely on links made using exact matches on all of the linking fields. For the OPTIMIZE-HF database, limiting the results to records linked on exact values of all linking fields would have missed substantial numbers of records containing data that differed only slightly between databases.

There are several reasons why a registry record might not exactly match a Medicare claim for the same hospitalization. Medicare may not know the sex of a beneficiary, in which case it assigns a value of "female."¹⁰ Moreover, a single Medicare claim may represent 2 distinct hospital stays if the patient was readmitted on the same day he was discharged for a diagnosis in the same diagnosis-related group.¹¹ In addition, the registry may indicate the date of arrival to the emergency department as the patient's admission date, which may differ from the admission date in the claims data. Finally, there may be data entry errors or other minor inconsistencies between databases on fields such as date of birth. Because of these slight differences, the correct link can be made without requiring fully deterministic matches.

There are many reasons why registry records may, appropriately, not link to any Medicare records. First, all medical care provided to patients who opt to enroll in a Medicare managed care plan is paid through a single capitated payment. Thus, no encounter-level bills are generated and these patients are not included in Medicare claims data. Second, patients receiving care at VA hospitals are not included in Medicare claims data, because VA hospitals are not paid through the Medicare system. Third, patients who have primary insurance coverage through private or employer-sponsored insurance plans will not appear in Medicare data if the primary insurance plan fully covers the cost of the inpatient stay. Finally, registry records that reflect outpatient hospital visits or claims for emergency department visits that do not result in an admission are not included in the inpatient Medicare data, because these claims are paid through the Medicare Part B outpatient benefit.

These potential reasons for mismatch create a ceiling on the proportion of registry hospitalizations that can be linked to Medicare inpatient claims data. This limitation will vary among registries, depending on the quality of the registry data, the proportion of registry records that actually represent outpatient visits, the number of VA hospitals participating in the registry, the proportion of patients who have other insurance coverage, and the Medicare managed care penetration among registry hospitals. Since 2000, Medicare managed care enrollment has accounted for an average of approximately 15% of all eligible Medicare beneficiaries, but geographic variation in managed care enrollment is substantial.¹² One reason for the higher proportion of linked records from the Duke database is the low Medicare managed care penetration in North Carolina. Based on our experience linking many different registries to Medicare data, we expect to be able to identify 70% to 80% of elderly registry patients in the Medicare claims data.

Linking data for longitudinal follow-up has many advantages. Obtaining this information from existing claims data is much less expensive than collecting follow-up data directly from patients or providers. Because the entire process is based on anonymous identifiers, the challenges of obtaining patient consent may potentially be waived. Therefore, this system would likely be

more complete than a system that requires hospitals and patients to opt in to follow-up. Thus, the resulting linked database will be less subject to participation-based selection biases that have posed challenges for other registries.

Although the use of Medicare data for longitudinal follow-up has several advantages, it also has limitations. Medicare claims exist only for patients aged 65 years or older and patients with qualified disability coverage. State-level all-payer claims files, Medicaid data, and major private insurer databases provide alternative options for linking patients younger than 65 years; however, each of these data sources covers only a select group of patients. In addition, using indirect identifiers to link databases may result in some low number of incorrect links. However, we have no reason to expect that incorrect links are systematic in ways that would produce bias in analyses based on these data, and other studies have validated the use of indirect identifiers to link health care databases.^{13,14}

Once registry patients have been identified in Medicare, we can characterize many types of postdischarge outcomes, including mortality, readmission, and subsequent inpatient procedures. Among the many types of research questions this data set allows, researchers can answer questions about long-term safety and efficacy of inpatient treatments and about the relative importance of hospital processes for long-term patient outcomes.

Conclusions

In the absence of direct identifiers, it is possible to create a high-quality link between inpatient clinical registry data and Medicare claims data. The method allows researchers to leverage existing data to create a linked claims-clinical database that capitalizes on the strengths of both types of data sources. Combined databases such as these are important at a time when there is otherwise little infrastructure to answer important safety, efficacy, and other clinical questions for large patient populations in real-world settings.

Steps for Linking Inpatient Registry Data to Medicare Inpatient Claims Data

1. Subset registry data
 - Limit to records for patients aged 65 years or older.
2. Subset Medicare data
 - Limit to records for patients aged 65 years or older.
 - Limit using broadly defined inclusion criteria to match registry entry criteria.
3. Link hospital identifiers
 - Link records between data sources on exact values of admission date, discharge date, patient date of birth or age, and patient sex.
 - Use resulting number of matches to inform links.
4. Determine specifications for hospitalization-level linking
 - Choose rules to apply.
 - Decide whether Medicare records should be allowed to link to multiple registry records.
5. Link hospitalization records

Acknowledgments

We thank Damon M. Seils, MA, Duke University, for assistance with manuscript preparation. Mr Seils did not receive compensation for his assistance apart from his employment at the institution where the study was conducted.

Funding/Support

This work was supported by grant U18HS10548 from the Agency for Healthcare Research and Quality and a research agreement between GlaxoSmithKline and Duke University. Dr Hernandez is a recipient of an American Heart Association Pharmaceutical Roundtable grant (0675060N). Drs Curtis and Schulman were supported in part by grants 5U01HL066461 from the National Heart, Lung, and Blood Institute and 1R01AG026038-01A1 from the National Institute on Aging. Dr Fonarow is supported by the Ahmanson Foundation and the Corday Family Foundation. The OPTIMIZE-HF registry is registered at clinicaltrials.gov as study number NCT00344513.

References

1. Fonarow GC, Abraham WT, Albert NM, et al. Influence of a performance-improvement initiative on quality of care for patients hospitalized with heart failure: results of the Organized Program to Initiate Lifesaving Treatment in Hospitalized Patients With Heart Failure (OPTIMIZE-HF). *Arch Intern Med* 2007;167:1493–502. [PubMed: 17646603]
2. Mehta RH, Roe MT, Chen AY, et al. Recent trends in the care of patients with non-ST-segment elevation acute coronary syndromes: insights from the CRUSADE initiative. *Arch Intern Med* 2006;166:2027–34. [PubMed: 17030838]
3. Peterson ED, Roe MT, Mulgund J, et al. Association between hospital process performance and outcomes among patients with acute coronary syndromes. *JAMA* 2006;295:1912–20. [PubMed: 16639050]
4. Tu JV, Willison DJ, Silver FL, et al. Impracticability of informed consent in the Registry of the Canadian Stroke Network. *N Engl J Med* 2004;350:1414–21. [PubMed: 15070791]
5. Fonarow GC, Abraham WT, Albert NM, et al. Organized Program to Initiate Lifesaving Treatment in Hospitalized Patients with Heart Failure (OPTIMIZE-HF): rationale and design. *Am Heart J* 2004;148:43–51. [PubMed: 15215791]
6. Bradley CJ, Given CW, Luo Z, et al. Medicaid, Medicare, and the Michigan Tumor Registry: a linkage strategy. *Med Decis Making* 2007;27:352–63. [PubMed: 17641138]
7. Lillard LA, Farmer MM. Linking Medicare and national survey data. *Ann Intern Med* 1997;127:691–5. [PubMed: 9382381]
8. Warren JL, Klabunde CN, Schrag D, et al. Overview of the SEER-Medicare data: content, research applications, and generalizability to the United States elderly population. *Med Care* 2002;40(8 Suppl):IV–18.
9. Herzog, TN.; Scheuren, FJ.; Winkler, WE. *Data Quality and Record Linkage Techniques*. Springer; New York: 2007.
10. Research Data Assistance Center. Record Layout for Medicare Denominator File. 2008. <http://www.resdac.umn.edu/ddde/SEX.txt> Accessed October 12, 2008
11. Centers for Medicare & Medicaid Services. Revision of Common Working File (CWF) Editing for Same-Day, Same-Provider Acute Care Readmissions. 2008. <http://www.cms.hhs.gov/MLNMattersArticles/downloads/MM3389.pdf> Accessed October 12, 2008
12. Centers for Medicare & Medicaid Services. Medicare Managed Care Contract Report (January 1, 2004). 2008. <http://www.cms.hhs.gov/HealthPlanRepFileData/Downloads/MMCC-2004.zip> Accessed October 12, 2008
13. Méray N, Reitsma JB, Ravelli AC, et al. Probabilistic record linkage is a valid and transparent tool to combine databases without a patient identification number. *J Clin Epidemiol* 2007;60:883–91. [PubMed: 17689804]
14. Newgard CD. Validation of probabilistic linkage to match de-identified ambulance records to a state trauma registry. *Acad Emerg Med* 2006;13:69–75. [PubMed: 16365326]

Table 1
Records in the 2004 Medicare Claims Data Unique for Specific Combinations of Variables Within Hospitals

Combination Rule	Variable						Unique Records, %	
	Admission Date	Discharge Date	Patient Sex	Patient Date of Birth	Patient Age	All Claims (N = 11.1 million)	Heart Failure Claims (n = 2.8 million)	
1 ^a	•	•	•	•	•	> 99.9	> 99.9	
2	•	•	•	•	•	> 99.9	> 99.9	
3	•	•	•	•	•	99.9	> 99.9	
4	•	•	•	•	•	99.9	> 99.9	
5	•	•	•	•	•	99.8	> 99.9	
6	•	•	•	•	•	99.8	99.9	
7 ^b	•	•	•	2 out of 3	•	99.4	99.9	
8	•	•	•	2 out of 3	•	98.9	99.8	
9	•	•	•	2 out of 3	•	95.3	98.8	
10	•	•	•	2 out of 3	•	95.1	98.7	
11	•	•	•	2 out of 3	•	91.1	97.7	
12	•	•	•	2 out of 3	•	90.9	97.5	
13	•	•	•	•	•	97.0	99.4	
14	•	•	•	•	•	94.4	98.9	
15 ^c	•	± 1 day	•	•	•	92.6	98.4	
16	± 1 day	•	•	•	•	92.5	98.4	
17 ^d	•	•	•	•	± 1 year	91.8	98.3	
18	•	± 1 day	•	•	•	86.7	97.0	
19	± 1 day	•	•	•	•	86.6	97.0	
20	•	•	•	•	± 1 year	85.5	96.8	
21	•	•	•	•	•	78.7	94.0	
22	•	•	•	•	•	78.2	93.5	
23	•	•	•	•	•	65.5	89.0	
24	•	•	•	•	•	64.7	88.2	

Example combinations:

^a A record is considered unique for this combination if no other record in the same hospital matches admission date, discharge date, sex, and date of birth.

- ^b A record is considered unique for this combination if no other record in the same hospital matches admission date, discharge date, sex, and at least 2 of 3 components of date of birth.
- ^c A record is considered unique for this combination if no other record in the same hospital matches admission date, discharge date within 1 day, sex, and age.
- ^d A record is considered unique for this combination if no other record in the same hospital matches admission date, discharge date, sex, and age within 1 year.

Table 2
Number of Inpatient Database Records Linked to Medicare Claims Data Using Each Combination Rule

Combination Rule ^a	No. (%)	
	OPTIMIZE-HF (N = 39,178)	Duke Hospital (N = 6581)
Rules using date of birth		
1	27,244 (69.5)	5573 (84.7)
2	277 (0.7)	6 (0.1)
3	1149 (2.9)	19 (0.3)
4	562 (1.4)	8 (0.1)
5	14 (< 0.1)	0
6	7 (< 0.1)	0
7	2230 (5.7)	402 (6.1)
8	22 (< 0.1)	0
9	109 (0.3)	5 (0.1)
10	136 (0.3)	0
Total	31,750 (81.0)	6013 (91.4)
Rules using age		
13	28,258 (72.1)	5830 (88.6)
14	322 (0.8)	8 (0.1)
15	313 (0.8)	4 (0.1)
16	863 (2.2)	2 (< 0.1)
17	935 (2.4)	130 (2.0)
Total	30,691 (78.3)	5974 (90.8)

^aFor each rule, the hospital identifiers were also required to match, using the established crosswalk. Rule numbers refer to the combinations described in Table 1.