Award:	**CERTs Comparative Effectiveness Supplement** (U18HS010399-06S1)
Project Title:	Validation of Outcomes in Centers for Medicare & Medicaid Services (CMS) Data
Principal Investigator:	Brian L. Strom, MD, MPH
Co-Principal Investigator:	Sean Hennessy, PharmD, PhD
Institution:	Center for Clinical Epidemiology & Biostatistics University of Pennsylvania School of Medicine Philadelphia, Pennsylvania
Date:	December 21, 2006

FINAL REPORT

We performed a methodologic study to demonstrate the feasibility of an academic-Clinical Data Abstraction Center (CDAC) partnership to verify hospitalization study outcomes using clinical records. The specific aims of this proposal were as follows:

1. Demonstrate the feasibility of obtaining approval and waiver of HIPAA authorization from an academic institutional review board to permit verification of outcomes without patient consent via an academic-CDAC partnership.
2. Demonstrate the feasibility of obtaining Data Use Agreements from CMS that permit validation of hospitalization outcomes, without patient consent, using an academic-CDAC partnership;
3. Demonstrate the feasibility of a CDAC obtaining, for research purposes, hospital records corresponding to CMS data likely to be used by researchers for studies of comparative effectiveness;
4. Identify barriers to the proposed model, strategies to overcome or avoid them, and potential improvements to the model;
5. Measure the positive predictive value of a specific outcome that might be employed in studies of comparative effectiveness.

Aims 1 through 4 are fully addressed in a manuscript currently under review by *Epidemiology* (Lippincott Williams & Wilkins, publisher). Please consider the enclosed manuscript as a component of our final report. In addition, the following addendum addresses the research conducted pursuant to Aim 5.

ADDENDUM

Objective: Investigators sought to measure the positive predictive value of a specific outcome that might be employed in studies of comparative effectiveness. In particular, we validated hospitalization for the composite outcome of sudden cardiac death and ventricular arrhythmia (SCD/VA) in an administrative dataset of 1999-2000 Medicaid and Medicare data using expert medical record review as the gold standard.

Background: Probably the most important concern with using claims data for research is its validity. A study funded by the US Food and Drug Administration and performed by the Research Triangle Institute in the early 1980s compared Medicaid encounter data from Michigan and Minnesota to its primary sources, i.e., clinical records hospitals, physician offices, pharmacies, etc.[1] The results of this study suggested that the demographic and drug data appeared to be of extremely high quality. Regarding medical services, of those records that could be evaluated, 93% of the services in Medicaid encounter data could be found in the provider records within one week of the Medicaid encounter date. However, in 17% of those, the provider record included a previous or subsequent visit that was not included in the Medicaid encounter data. Diagnostic agreement to at least 3 digits of the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9) code occurred in 41%, while agreement within a broad diagnostic category in another 16% (i.e., same body system and/or type of illness). No diagnosis was present on the provider record in 12%, and

there was no agreement in 28%. These results suggest that the validity of diagnostic data must be considered in the context of each individual study outcome.

The validity of a number of specific Medicaid encounter diagnoses have been examined using primary medical records, with each study illustrating strengths and weaknesses of Medicaid data. In a study by Brian Strom, MD, MPH of the validity of neutropenia diagnoses in Medicaid, the encounter diagnosis was verified by laboratory information in 192 of 198 clinical records available, yielding a positive predictive value of 97%.[2] However, the purpose of the study was to investigate incident cases of neutropenia, yet 13.5% of the cases had recurrent neutropenia, and 9.9% had cyclic neutropenia. Thus, even when the encounter diagnosis was highly accurate, other information from the medical record was needed to characterize cases accurately. For a Medicaid study of Stevens-Johnson Syndrome also performed by Dr. Strom, the hospital records of 249 cases with an inpatient diagnosis of erythema multiforme (ICD-9 code 695.1) were sought from three states.[3] Of these, 128 (51.4%) medical records were available. Of these, 121 (94.5%) subjects had received a clinical diagnosis that was potentially compatible with the encounter diagnosis. However, upon review by a study dermatologist, a diagnosis of erythema multiforme minor or major was confirmed in only 42% of those with the relevant encounter diagnosis. This study confirmed that this ICD-9 diagnosis code includes several unrelated conditions, and that the clinical conditions are not always accurately coded. Based on these and other studies, our experience suggests that in any study using claims data, with few exceptions, investigators should obtain medical records in at least a sample of outcomes to confirm the validity of the encounter diagnoses, characterize the severity of the disease, and obtain information on potential confounding variables not found in the encounter data.

Sudden cardiac death (SCD) is defined as the sudden, abrupt loss of heart function.[4] It is considered a major cause of mortality, and a major public health concern. Only about 5% of SCD cases survive.[4] SCD results in 300,000-456,000 deaths per year, or about 12-19% of all deaths in the US,[5,6] and is the leading cause of death in adults less than 65 years old.[7] SCD accounts for 50-63% of all cardiovascular deaths,[5,6] with over half of SCD cases having no prior diagnosis of heart disease.[8] The annual incidence of SCD is 1.8 per 1000 in the general US population,[9] which mirrors the rate of about 2 per 1000 per year in Medicaid recipients receiving prescription medications.[10,11] Ventricular fibrillation (VF) is a pulseless arrhythmia with irregular and chaotic electrical activity and ventricular contraction in which the heart immediately loses its ability to function as a pump.[12] VF is the initial electrocardiogram (ECG) rhythm in 75% of outpatient cases of SCD.[13] Ventricular tachycardia (VT) is a rhythm >100 beats per minute arising from the bundle of His. Because of this close relationship between S**C**D and VA (VF and VT), some authors prefer the term "presumed arrhythmic death" to SCD.[14] Validation of this outcome in administrative data has been limited. A study by Staffa et al, using a slightly less restrictive validation definition, validated 8 of 11(73%) outcomes via medical record review when compared to Ohio Medicaid claims data.[15]

Research design and methods

Study design: Investigators conducted manual medical record reviews of retrospectively-identified putative cases of SCD/VA in 1999 and 2000 Medicaid and Medicare data from CA, FL, NY, OH, and PA. These states constitute approximately 30 million person-years of Centers for Medicare & Medicaid Services (CMS) data.

Study site: Inpatient medical records were obtained, by the investigators, from the Computer Sciences Corporation (CSC) located in York, Pennsylvania. CSC had functioned as a quality assurance agent for CMS, as the national CDAC, requesting medical records from hospitals across the country on CMS' behalf. The investigators leveraged CSC's experience in this realm to have them request records of interest from hospitals in the aforementioned states. These records, redacted of direct personal identifiers, were then shipped by CSC to the investigators for review. Please refer to our manuscript under review by *Epidemiology* for further details of this process.

Case identification: Utilizing CMS data, investigators identified inpatient diagnoses of SCD/VA via presence of an ICD-9 claims diagnosis. Table 1 includes diagnosis codes of interest and was developed in concert with an expert cardiologist with attention to codes utilized in similar studies. In particular, diagnoses were identified in the inpatient Medicaid Analytic Extract (MAX IP) and Medicare Provider and Analysis Review (MEDPAR) files. Investigators then selected a random sample of events and forwarded to CSC to initiate the retrieval process.

Table1. SCD/VA ICD-9 diagnosis codes

ICD-9 Code	Description
427.1	Paroxysmal Ventricular Tachycardia; Ventricular tachycardia (paroxysmal)
427.4	Ventricular Fibrillation And Flutter
427.41	Ventricular Fibrillation

427.42	Ventricular Flutter
427.5	Cardiac Arrest; Cardiorespiratory arrest
798	Sudden Death, Cause Unknown
798.1	Instantaneous Death
798.2	Death Occurring In Less Than 24 Hours From Onset Of Symptoms, Not Otherwise Explained; Death known not to be violent or instantaneous, for which no cause could be resolved; Died without sign of disease

Validation Definition: A priori, the investigators defined the composite outcome of SCD/VA as follows, as adapted from that used by Ray,[16]: an outpatient witnessed sudden collapse, person found unconscious or dead with evidence that the person had been alive in the preceding 24 hours, or evidenced/witnessed cardiac arrest or ventricular arrhythmia. The definition excluded episodes originating in the hospital, but included those originating in a nursing home. Events with documentation suggesting an extrinsic (e.g., injury) or a non-cardiac cause (e.g., pneumonia) were excluded. Documentation on ECG was not required to meet the definition, but was recorded. As medical record validation was deemed the gold standard, this definition was used during the review process to validate each putative event.

Abstraction Process: Of the randomly-selected medical records received from CSC, a specially-trained clinical pharmacist and public health researcher teamed to abstract each medical record onto an electronic abstraction form. This form was reviewed and approved by the Co-Principal Investigator (co-PI) and a cardiologist. Abstractors utilized a reproducible algorithm to score each abstraction form as meeting or failing to meet the validation definition. Those deemed unclear were secondarily reviewed by the co-PI who acted as the final decision-maker.

Statistical Analysis: A validity measure was calculated for the ICD-9 codeset identified in Table 1 above. Presence of a claims diagnosis from the codeset was considered a positive "test" for SCD/VA. The abstractors'/co-PI's determination of validity (via medical record review) was considered the gold standard. Positive predictive value (PPV) was defined as the probability that the gold standard determined positivity given that the "test" was positive. Binomial exact confidence intervals for proportions were calculated using Stata v9.1 (Stata Corp, College Station, TX: 2005).

Results: Queries of the CMS claims data identified 477 SCD/VA events occurring within 127,106 person-years while beneficiaries were exposed to either cisapride, metoclopramide, or a proton pump inhibitor. As this study was conducted supplementary to an ongoing R01 investigating drug-induced SCD/VA, this nested cohort had already been identified. 164 medical records were randomly-selected by the investigators and provided to CSC, of which 128 (78%) were received and redacted. All 128 medical records were abstracted by either the clinical pharmacist (N=59, 46%) or public health researcher (N=69, 54%). Ten medical records (7.8%) underwent secondary review by the co-PI (trained as a clinical pharmacist and epidemiologist). Of the 128 records, 23 (18%, 95%CI [12-26%]) met the validation definition while 105 (82%, [74-88%]) did not. A subanalysis identified seven medical records with a SCD/VA ICD-9 diagnosis listed as the primary claim diagnosis. Seven of seven (100%, [65-100%]) met the validation definition.

Discussion: The positive predictive value (PPV) of SCD/VA diagnosis codes of 18% was much lower than anticipated given results from a previous study[15] and pilot work conducted within our institution. Our result, though, was confirmed by a similar validation study we have conducted within the General Practice Research Database (PPV=17%) (data not presented). Our validation identified that the vast majority of SCD/VA events did occur, albeit while in-hospital, most considerably post-presentation. If we had not restricted our validation definition to exclude events occurring during the course of hospitalization, the PPV would have been significantly greater. In our defense, the validation definition developed for this study was most appropriate for the study of outpatient SCD/VA (intent of the parent R01) and was consistent with that used by other researchers.[16] The analysis of a subset of events with a SCD/VA code as the primary claims diagnosis demonstrated a much different result. While the sample was sufficiently small (N=7), the PPV is at least 65% and consistent up through and including 100%. This potentially argues that studies of utilizing this outcome should include only those SCD/VA events indicated as primary diagnoses. Alternatively, researchers wishing to use all available inpatient diagnoses should make efforts to validate all putative cases and include only those meeting their validation definition.

Conclusion: Overall, our outpatient-occurring composite SCD/VA outcome validated poorly (PPV= 18%). This may be improved by focusing on primary claims diagnoses only or utilizing, as true cases, only those that can be validated via medical record review.

Reference List

1. Lessler, J. T. & Harris, B. S. H. Medicaid data as a source for postmarketing surveillance information. 1984. Research Triangle Park, NC, Research Triangle Institute.
   Ref Type: Report

2. Strom, B. L., Carson, J. L., Schinnar, R., Snyder, E. S. & Shaw, M. Descriptive epidemiology of agranulocytosis. *Arch. Intern. Med.* **152**, 1475-1480 (1992).

3. Strom, B. L. *et al.* A population-based study of Stevens-Johnson syndrome. Incidence and antecedent drug exposures. *Arch. Dermatol.* **127**, 831-838 (1991).

4. American Heart Association. Sudden deaths from cardiac arrest. 2003.
   Ref Type: Electronic Citation

5. Saliba, W. I. & Natale, A. Ventricular tachycardia syndromes. *Med. Clin. North Am.* **85**, 267-304 (2001).

6. Zheng, Z. J., Croft, J. B., Giles, W. H. & Mensah, G. A. Sudden cardiac death in the United States, 1989 to 1998. *Circulation* **104**, 2158-2163 (2001).

7. Cupples, L. A., Gagnon, D. R. & Kannel, W. B. Long- and short-term risk of sudden coronary death. *Circulation* **85**, I11-I18 (1992).

8. Kannel, W. B., Wilson, P. W., D'Agostino, R. B. & Cobb, J. Sudden coronary death in women. *Am. Heart J.* **136**, 205-212 (1998).

9. State-specific mortality from sudden cardiac death--United States, 1999. *MMWR Morb. Mortal. Wkly. Rep.* **51**, 123-126 (2002).

10. Pratt, C. M. *et al.* Risk of developing life-threatening ventricular arrhythmia associated with tefenadine in comparison with over-the-counter antihistamines, ibuprofen and clemastine. *Am. J. Cardiol.* **73**, 346-352 (1994).

11. Hennessy, S. *et al.* Cardiac arrest and ventricular arrhythmia in patients taking antipsychotic drugs: cohort study using administrative data. *BMJ.* **325**, 1070 (2002).

12. Oliver, M. F. Metabolic causes and prevention of ventricular fibrillation during acute coronary syndromes. *Am. J. Med.* **112**, 305-311 (2002).

13. Greene, H. L. Sudden arrhythmic cardiac death--mechanisms, resuscitation and classification: the Seattle perspective. *Am. J. Cardiol.* **65**, 4B-12B (1990).

14. Greene, H. L. *et al.* Classification of deaths after myocardial infarction as arrhythmic or nonarrhythmic (the Cardiac Arrhythmia Pilot Study). *Am. J. Cardiol.* **63**, 1-6 (1989).

15. Staffa, J. A., Jones, J. K., Gable, C. B., Verspeelt, J. P. & Amery, W. K. Risk of selected serious cardiac events among new users of antihistamines. *Clin. Ther.* **17**, 1062-1077 (1995).

16. Ray, W. A. *et al.* Antipsychotics and the risk of sudden cardiac death. *Arch. Gen. Psychiatry* **58**, 1161-1167 (2001).