Sensitivity of Billing Claims for Cardiovascular Disease Events among Kidney Transplant Recipients

Krista L. Lentine,* Mark A. Schnitzler,* Kevin C. Abbott,† Kosha Bramesfeld,‡ Paula M. Buchanan,* and Daniel C. Brennan§

*Center for Outcomes Research and Division of Nephrology, School of Medicine, and ‡Department of Research Methodology, Graduate School, Saint Louis University, and §Division of Nephrology, Washington University School of Medicine, St. Louis, Missouri; and †Nephrology Service, Walter Reed Army Medical Center, Washington, DC

Background and objectives: Billing claims are increasingly examined beyond administrative functions as outcomes measures in observational research. Few studies have described the performance of billing claims as surrogate measures of clinical events among kidney transplant recipients.

Design, setting, participants, & measurements: We investigated the sensitivity of Medicare billing claims for clinically verified cardiovascular diagnoses (five categories) and procedures (four categories) in a novel database linking Medicare claims to electronic medical records of one transplant program. Cardiovascular events identified in medical records for 571 Medicare-insured transplant recipients in 1991 through 2002 served as reference measures.

Results: Within a claims-ascertainment period spanning ±30 d of clinically recorded dates, aggregate sensitivity of single claims was higher for case definitions incorporating Medicare Parts A and B for diagnoses and procedures (90.9%) compared with either Part A (82.3%) or Part B (84.6%) alone. Perfect capture of the four procedures was possible within ±30 d or with short claims window expansion, but sensitivity for the diagnoses trended lower with all study algorithms (91.2% with window up to ±90 d). Requirement for additional confirmatory diagnosis claims did not appreciably reduce sensitivity. Sensitivity patterns were similar in the early compared with late periods of the study.

Conclusions: Combined use of Medicare Parts A and B billing claims composes a sensitive measure of cardiovascular events after kidney transplant. Further research is needed to define algorithms that maximize specificity as well as sensitivity of claims from Medicare and other insurers as research measures in this population.

Cardiovascular events account for 30 to 50% of mortality among kidney transplant recipients and compose one of the most important public health concerns in this population (1,2). Single-center medical records may provide detailed descriptions of clinical complications such as cardiovascular diagnoses and procedures after transplantation, but medical records are difficult to analyze unless integrated and stored electronically. The Organ Procurement and Transplantation Network (OPTN) collects national survey data on comorbidities at transplant but not in follow-up forms, and the current strategic focus of the OPTN is to reduce rather than broaden the scope of survey data collection to lower time and effort burdens on centers (3,4). Administrative billing claims of insurance providers compile dated, diagnosis-linked records of physician encounters, hospitalizations, and procedures that may be useful as proxy measures of clinical cardiovascular events in epidemiologic studies of large samples, including populations of transplant recipients.

The *International Classification of Diseases, Ninth Revision, Clinical Modification* (ICD-9-CM) is an internationally accepted method for classifying diseases, diagnoses, and procedures in the processing of bills for medical services (5). The ICD-9-CM codes on claims submitted to national insurance providers such as Medicare are increasingly examined beyond their administrative functions as surrogate measures in studies of health care use, care quality, clinical outcomes, and syndromic surveillance (6–8). Despite the growing popularity of billing claims as epidemiologic research measures, there is a relative paucity of data validating the accuracy of claims in representing clinical diagnoses and exposures in general population and transplantation samples. Among the published claims validation studies, Hebert *et al.* (9) studied the ability of Medicare claims with diagnosis codes for diabetes to identify self-reported diabetes among 6958 elderly, rural adults surveyed in 1992 through 1993. The authors found that a minimum requirement of one institutional claim or two ambulatory claims in the previous 2 yr provided good accuracy for research purposes ($\kappa = 0.77$, sensitivity 0.76, specificity 0.98).

The "Hebert case definition" caught the attention of epidemiologists in transplantation, and related algorithms have been applied to describe not only the incidence of posttransplantation diabetes but also an array of clinical conditions in kidney transplant recipients, including malignant, infectious, gastrointestinal, metabolic, psychiatric, and some cardiovascular complications (10–18). Other studies of acute posttransplantation cardiovascular events have used single claims as case definitions (19–21). On the basis of Medicare data for elderly Pennsylvania residents in 1999 through 2000, Kiyota et al. (22) found that single-hospital claims with diagnosis codes for acute myocardial infarction offered 94% positive predictive value (a metric of measurement accuracy that frames specificity in the context of prevalence) for myocardial infarctions recorded in hospital records, but these authors did not report sensitivity metrics for the claims-based definition. The accuracy of Medicare claims for cardiovascular events has not been assessed for patients with insurance eligibility on the basis of transplantation rather than age.

To improve understanding of the accuracy and limitations of billing claims as measures of cardiovascular events among kidney transplant recipients, we examined a novel, aggregate database of electronic medical records from one large Midwestern transplant program linked at the patient level with administrative billing claims of Medicare. Using this compilation of two independent data sources, we assessed the sensitivity of Medicare billing claims to detect clinically verified cardiovascular events and investigated whether sensitivity differed among several claims algorithms. In particular, we examined (1) sensitivity of combined Part A and Part B claims, as compared with use of either billing source alone; (2) comprehensiveness of detection in claims for procedures compared with diagnoses; (3) possible reductions in sensitivity for diagnoses with application of a more stringent requirement for additional confirmatory claims in the case definition; (4) the impact of extended claims capture periods on sensitivity; and (5) sensitivity patterns over time during the study.

Materials and Methods
Data Sources and Linkage Methods
The data source for this study is a novel construction in which electronic medical records from the Washington University Kidney Transplant Program Database (WU-KTDB) were linked at the patient level with national OPTN records and Medicare billing claims as compiled within the US Renal Data System (USRDS). After institutional review board, National Institutes of Health, and USRDS approvals, recipient identifier numbers from the WU-KTDB were linked using Social Security Numbers, names, and dates of birth to USRDS_IDs. Names and Social Security Numbers were removed after the linkage procedure, such that records for individual patients in the analytic files are identified by anonymous WU-KTDB patient codes and USRDS_IDs. Match validation included confirmation of agreement of transplant dates recorded in clinical records with OPTN reports.

Sampling Criteria
For this analysis, we retrospectively sampled adult (≥18 yr of age) recipients of kidney transplants at Washington University in 1991 through 2002. To allow examination of Medicare claims as diagnosis

and procedure measures, analytic samples were restricted to patients with Medicare as their primary insurer. Medicare payer status for transplant recipients is tracked within "Payer History" records of the USRDS (23); we also require a minimum Medicare payment for the initial transplant hospitalization of at least $15,000 as indication of usage of Medicare primary benefits (24). Patients included in measures agreement analyses for particular clinical events were required to have active Medicare benefits at the time of the reference event, and thus analytic subsamples varied according to the event of interest.

Measures of Cardiovascular Diagnoses and Procedures
The American Society of Transplantation defines cardiovascular disease as coronary heart disease, cerebrovascular disease, and peripheral vascular disease (25). We also studied congestive heart failure, atrial fibrillation, and venous thromboembolism as cardiovascular diagnoses that have been associated with morbidity and mortality after transplantation (12,15,16,26). Along with diagnoses, we studied several cardiovascular procedures tracked in the WU-KTDB: Cardiac catheterization, coronary artery bypass grafting, amputation, and revascularization of peripheral vascular disease.

Reference Standard for Events
The clinical criteria used for scoring cardiovascular events are defined in Table 1. Information describing the clinical course of Washington University's transplant recipients is prospectively and retrospectively entered into a secured research database by trained nurse coordinators. Out-of-center events are included when they are reported back to the center with sufficient supporting detail to meet diagnostic criteria. Because many transplant recipients return to community providers for general care and some outside events are not communicated to the center, the transplant center data were not expected to capture comprehensively all out-of-center cardiovascular events but were considered as the reference standard for clinically verified events when captured. Complete follow-up of patient vital status and allograft status is maintained for all transplant recipients at the center.

Billing Claims as Comparison Measures
Evidence of clinical events was ascertained from Medicare billing claims using ICD-9-CM Diagnosis Codes, ICD-9 Procedure Codes, and Current Procedural Terminology codes submitted with claims. Table 2 lists the specific codes used to identify the cardiovascular diagnoses and procedures of interest within claims-based algorithms.

Statistical Analysis
We explored a series of algorithms for claims-based ascertainment of diagnoses and procedures. First, we examined case definitions formed by single institutional claims (Medicare Part A), single physician/supplier claims (Medicare Part B), or one claim of either type (Part A or B) within ±30 d of event date recorded in the WU-KTDB. This primary capture window was chosen because dates of some events in the clinical record were recorded as month and year, without day of month. Next, we examined case capture by expanding the window of eligible single claims successively to within ±45, ±60, and ±90 d of the clinically recorded event.

For diagnoses only, we also applied a variation of the Hebert method similar to that used by the USRDS Annual Data Report, which defines diagnoses on the basis of one Part A or two Part B claims submitted at least 1 d but no more than 365 d apart, in which the latest claim date is defined as date of diagnosis (2); we allowed final diagnosis date to fall up to 30 d after the date in the clinical record. In contrast to diagnoses that may be confirmed or excluded in practice on the basis of clinical reassessment,

Table 1. Definitions of cardiovascular diagnoses and procedures as recorded in the transplant center's clinical database

Parameter	Clinical Criteria
Cardiovascular diagnoses	
myocardial infarction	"Definite" or "probable" events according to the Minnesota code electrocardiographic and biomarker criteria adopted by the American Heart Association for use in clinical research (30).
congestive heart failure	Guidelines incorporate into the diagnosis of heart failure clinical judgment on the basis of history and physical examination (31). Accordingly, indication of congestive heart failure within the database requires physician-reported diagnosis plus objective evidence of cardiac dysfunction (*e.g.*, echocardiography or other forms of ventriculography, chest radiograph, and/or B-natriuretic peptide).
atrial fibrillation	Atrial fibrillation pattern on electrocardiography.
stroke	New focal neurologic deficit lasting ≥24 h, confirmed by brain imaging (computed tomography or magnetic resonance)
transient ischemic attack	New focal deficit that resolves within 24 h and was attributed to a central cause by the examining provider.
venous thromboembolism	Deep venous thrombosis was defined by clinical suspicion with confirmation by compression ultrasound of an upper or lower extremity. Pulmonary embolism was defined as positive or "high probability" findings by ventilation-perfusion lung scanning, spiral computed tomographic angiography of the chest, or pulmonary angiography.
Procedures	
cardiac catheterization	Coronary angiography performed for diagnostic or therapeutic purposes
coronary artery bypass grafting	Surgical revascularization of one or more coronary arteries
amputation	Surgical amputation of any portion of an upper or lower extremity
peripheral arterial revascularization	Angioplasty, atherectomy, or endarterectomy of one or more peripheral arteries or surgical arterial bypass

procedures are discrete events that may be represented by single claims; therefore, Hebert case definitions were not applied to procedures.

A schematic of the study design and event ascertainment procedures from the two data sources is shown in Figure 1. Data sets were merged and analyzed with SAS 9.1 for Windows (SAS Institute, Cary, NC). The sensitivity of each claims-based algorithm was computed as the proportion of reference events in the WU-KTDB among Medicare-insured patients that were captured by claims. Confidence intervals (CI) were computed as 95% CI for corresponding proportions. Differences in proportions were compared by two-proportions Z tests. The test statistic is computed as follows: $Z = (p_{1\_hat} - p_{2\_hat})/\{[p_{p\_hat}(1 - p_{p\_hat})]^{1/2} *[(1/n_1) + (1/n_2)]^{1/2}\}$, wherein the proportions compared are $p_{1\_hat} = (x_1/n_1)$ and $p_{2\_hat} = (x_2/n_2)$, and $p_{p\_hat} = (x_1 + x_2/n_1 + n_2)$. Hypotheses examined included whether (*1*) sensitivity of combined Part A or B claims (±30 d) as a case definition across all diagnoses and procedures was superior to sensitivity achieved with use of Part A or Part B claims alone; (*2*) claims sensitivity varied according to event type (*i.e.*, differed for procedures *versus* diagnoses); (*3*) claims sensitivity improved with capture window expansion; (*4*) the requirement for additional confirmatory diagnosis claims (Hebert approach) imposed a

cost of reduced sensitivity; and (*5*) Sensitivity patterns changed over time during the study (1991 through 1997 *versus* 1998 through 2002). Hypotheses with a rational directionality were examined with a one-tailed test. Specifically, the addition of claims data for event ascertainment by inclusion of both Medicare parts or by an expanded capture window could increase sensitivity or produce no change but not reduce sensitivity (one-tailed); the requirement for additional confirmatory diagnosis claims could reduce or not alter the sensitivity achieved with single claims but not increase sensitivity (one-tailed). Differences in sensitivity between procedures and diagnoses and across time periods of the study were examined with a more conservative two-tailed test because of the possibility of superiority for either event type or of either time period.

Results

A total 1128 kidney transplants recipients were recorded in the WU-KTDB in 1991 through 2002, and all matched uniquely to USRDS_IDs; of these, 571 met criteria for primary Medicare insurance at transplantation. The number of individual cardio-

Table 2. Billing claim codes used for identification of cardiovascular diagnoses and procedures[a]

Parameter	Billing Claim Codes
Cardiovascular diagnoses	
myocardial infarction	ICD-9-CM diagnosis: 410.x
congestive heart failure	ICD-9-CM diagnosis: 398.91, 422, 425, 428, 402.x1, 404.x1, 404.x3, V42.1
atrial fibrillation	ICD-9-CM diagnosis: 427.31x
stroke	ICD-9-CM diagnosis: 430, 431, 432, 433.x1, 434.x1, 997.02
transient ischemic attack	ICD-9-CM diagnosis: 435.x
venous thromboembolism	ICD-9-CM diagnosis: 453.4x, 453.8, 453.9, 415.11, 415.19
Procedures	
cardiac catheterization	ICD-9-CM procedure: 37.21, 37.22, 37.23, 88.53 to 88.57 CPT: 93508, 93510, 93511, 93514, 93524, 93526, 93527, 93528, 93529, 93530, 93531 to 93533, 93539 to 93545, 93555, 93556, 93598
coronary artery bypass grafting	ICD-9-CM procedure: 36.1x, 36.2 CPT: 33510 to 33523, 33533 to 33536
amputation	ICD-9-CM procedure: 84.0x, 84.1x, 84.91; CPT: 24900, 24920, 25900, 25905, 25920, 25927, 27295, 27590, 27591, 27592, 27598, 27880, 27881, 27882, 27888, 27889, 28800, 28805
peripheral arterial revascularization: angioplasty, atherectomy, endarterectomy, or arterial bypass	ICD-9-CM procedure: 39.25, 39.26, 39.29 CPT: 35331, 35341, 35351, 35355, 35361, 35363, 35371, 35372, 35381, 35450, 35452, 35454, 35456, 35459, 35470, 35471, 35472, 35473, 35474, 35480 to 35483, 35485, 35490 to 35493, 35495, 35521, 35531, 35533, 35541, 35546, 35548, 35549, 35551, 35556, 35558, 35563, 35565, 35566, 35571, 35583, 35585, 35587, 35621, 35623, 35646, 35647, 35651, 35654, 35656, 35661, 35663, 35665, 35666, 35671

[a]".x" indicates all fourth- and fifth-digit modifiers, unless otherwise noted. CPT, Common Procedural Terminology code; ICD-9-CM, *International Classification of Diseases, Ninth Revision, Clinical Modification.*
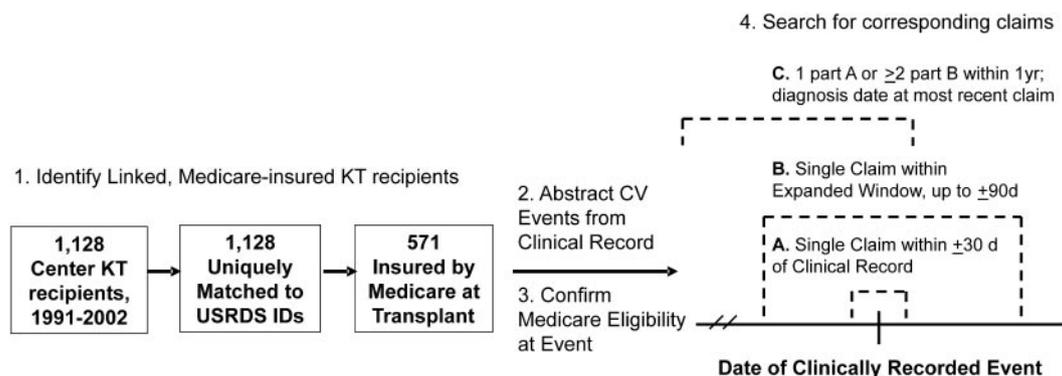

Figure 1. Study design: Sampling, clinical event abstraction, and billing claims ascertainment. CV, cardiovascular; KT, kidney transplant; A, primary claims capture window; B, expanded window for claims ascertainment; C, adapted-Hebert algorithm for detection of diagnoses in claims.

vascular events per diagnosis or procedure category in the clinical data ranged from 19 to 85; event counts during periods of reported Medicare eligibility ranged from 4 to 40 (Table 3).

Within a ±30 d window for claims ascertainment, the range of sensitivities of claims for clinically recorded cardiovascular events was 75.0 to 100.0% for Part A claims alone and 75.0 to

100.0% for Part B claims alone but improved to 83.3 to 100.0% for case definitions using Part A or B claims. The improvement with combined use of information from both billing sources compared with either source alone was statistically significant, with aggregate sensitivity of single Parts A or B claims across diagnoses and procedures of 90.9% compared with 82.3% for

Table 3. Sensitivity of Medicare claims for detection of clinically recorded cardiovascular diagnoses and procedures, according to claims-based ascertainment algorithm[a]

Event	No. in Clinical Database (Reference)		Single claims, ±30-d Window for Claims Ascertainment (Sensitivity % [95% CI])			Single Claims, Expanded Window for Claims Ascertainment (Sensitivity % [95% CI])			Adapted "Hebert Method" for Diagnoses (Sensitivity % [95% CI])[c]
	Total	Medicare Eligible[b]	Part A	Part B	Part A or B	Part A	Part B	Part A or B	
Diagnoses									
myocardial infarction	82	30	80.0% (78.0 to 92.0%)	83.3% (72.1 to 94.5%)	83.3% (72.1 to 94.5%)	NC	NC	NC	80.0% (78.0 to 92.0%)
congestive heart failure exacerbation	85	40	75.0% (63.7 to 86.3%)	85.0% (75.7 to 94.3%)	92.5% (85.6 to 99.4%)	77.5% (66.5 to 86.4%)[e]	NC	NC	92.5% (85.6 to 99.4%)
atrial fibrillation	51	32	81.3% (67.7 to 94.8%)	84.4% (71.8 to 97.0%)	93.8% (85.4 to 100.0%)	90.6% (80.5 to 100.0%)[f]	87.5% (76.0 to 99.0%)[e]	96.9% (90.9 to 100.0%)[e]	93.8% (85.4 to 100.0%)
stroke or transient ischemic attack	27	16	75.0% (53.8 to 96.2%)	81.3% (62.1 to 100.0%)	87.5% (71.3 to 100.0%)	NC	NC	NC	87.5% (71.3 to 100.0%)
venous thromboembolism	38	18	83.3% (66.1 to 100.0%)	77.8% (58.6 to 97.0%)	83.3% (66.1 to 100.0%)	88.9% (74.4 to 100.0%)[f]	88.9% (74.4 to 100.0%)[f]	94.4% (83.3 to 100.0%)[f]	83.3% (66.1 to 100.0%)
Procedures									
cardiac catheterization	64	22	95.4% (66.7 to 100.0%)	90.9% (78.9 to 100.0%)	95.4% (66.7 to 100.0%)	100.0%[d]	95.4% (66.7 to 100.0%)[e]	100.0%[d]	–
coronary artery bypass grafting	19	4	100.0%	75.0% (32.6 to 100.0%)	100.0%	NC	NC	NC	–
amputation	25	8	100.0%	87.5% (64.6 to 100.0%)	100.0%	NC	NC	NC	–
peripheral revascularization	21	6	83.3% (53.5 to 100.0%)	100.0%	100.0%	NC	NC	NC	–

[a]CI, confidence interval.

[b]Claims sensitivity was computed using a reference standard of events recorded in the clinical database among Medicare-insured transplant recipients with sustained Medicare benefits at the time of the clinical event of interest.

[c]Diagnoses based on one Part A claim or two Part B claims submitted at least 1 d but no more than 365 d apart, in which case the latest claim date is defined as date of diagnosis; final diagnosis date was allowed to fall up to 30 d after the date in the clinical record. Medicare Part A represents institutional claims; Medicare Part B represents physician/suppliers.

Expanded claims-ascertainment windows were explored at [d]±45, [e]±60, and [f]±90 d. Results for the narrowest expanded window that maximized detection of events in the clinical record are reported. NC, no change in sensitivity with expanded window up to ±90 d.

Part A (Z = 2.35, one-tailed P = 0.009) and 84.6% for Part B (Z = 1.79, one-tailed P = 0.037) alone (Table 4).

Perfect capture of the four studied procedures by claims was possible within ±30 d or with short claims window expansion. The sensitivity of combined Part A or B claims sought within ±30 d of clinically record dates was 100% for coronary bypass grafting, amputation, and peripheral revascularization. Capture of cardiac catheterization by Part A or B was nearly complete (95.4%) within ±30 d and was perfect (100.0%) with lengthening the claims capture window to ±45 d. In contrast, sensitivity of claims for the studied diagnoses trended lower with all study algorithms (91.2% with window up to ±90 d; Z = 1.95, two-tailed P = 0.05).

Overall, expansion of the period for claims ascertainment up to ±90 d did not affect sensitivity of combined Part A or B claims for the collection of study events (P = 0.22). Secondary analyses suggested higher capture of potentially chronic diagnoses of atrial fibrillation and venous thromboembolism with window expansion, in that sensitivity of the combined-claims algorithm for atrial fibrillation was maximized at 96.9% by window expansion to ±60 d and sensitivity for venous thromboemoblism was maximized at 94.4% by window expansion to ±90 d. Aggregate sensitivity with extended capture window up to ±90 d showed a nonsignificant trend toward higher capture than that achieved by querying claims within ±30 d of the clinically recorded date for these two diagnoses (96.0 *versus* 90.0%; Z = 1.18, one-tailed P = 0.12). Sensitivity was unchanged for myocardial infarction, heart failure, or stroke, and the pooled sensitivity of claims in a ±90-d window for events aside from atrial fibrillation and venous thromboembolism was of similar magnitude to that achieved within the primary capture window (92.1 *versus* 91.3%; Z = 0.22, one-tailed P = 0.41).

The requirement for additional confirmatory diagnosis claims in an adapted Hebert algorithm produced a slight decrease in maximal sensitivity achieved by single Part A or B claims for myocardial infarction but otherwise did not reduce sensitivity for the studied cardiovascular diagnoses. The overall detection of diagnoses with the adapted Hebert algorithm was of similar magnitude and statistically indistinguishable from that achieved with single Part A or B claims (88.2 *versus* 89.0%; Z = −0.19, one-tailed P = 0.43).

Sensitivity patterns were similar across the years of the study. Overall, there were 75 eligible cardiovascular events (57 diagnoses, 18 procedures) in the reference clinical data in 1991 through 1997 and 101 events (79 diagnoses, 22 procedures) in 1998 through 2002. The sensitivity of claims within a ±30 d window for study events in 1991 through 1997 compared with 1998 through 2002 was as follows: All events, 92.0 *versus* 90.1% (Z = 0.43, two-tailed P = 0.67); diagnoses, 91.2 *versus* 87.3% (Z = 0.71, two-tailed P = 0.48); and procedures, 94.4 *versus* 100% (Z = −1.12, two-tailed P = 0.27).

Discussion

Reductions in survey data collection directed at lowering time and effort burdens on transplant centers has been deemed a key focus of the OPTN strategic plan (3,4). In this policy

Table 4. Statistical comparison of differences in the overall sensitivity of claims according to case definition and event type

Comparison	Aggregate Sensitivity, $p1_{hat}$ (x_1/n_1)	Aggregate Sensitivity, $p2_{hat}$ (x_2/n_2)	Z Statistic[a]	P
Part A or B (1) *versus* Part A alone (2), all events, ±30-d ascertainment window	90.9% (160/176)	82.4% (145/176)	2.35	0.009[b]
Part A or B (1) *versus* Part B alone (2), all events, ±30-d ascertainment window	90.9% (160/176)	84.6% (149/176)	1.79	0.037[b]
Procedures (1) *versus* diagnoses (2), Part A or B claims, up to ±90-d ascertainment window	100% (40/40)	91.2% (124/136)	1.95	0.05[c]
Claims ascertainment window up to ±90 d (1) *versus* ±30 d (2), using Part A or B claims				
all events	93.2% (164/176)	90.9% (160/176)	0.79	0.22[b]
atrial fibrillation and venous thromboembolism	96.0% (48/50)	90.0% (45/50)	1.18	0.12[b]
all other events	92.1% (116/126)	91.3% (115/126)	0.22	0.41[b]
Adapted "Hebert method" (1) *versus* Part A or B claims (2) for diagnoses	88.2% (120/136)	89.0% (121/136)	−0.19	0.43[b]

[a]Z = $(p_{1-hat} - p_{2-hat})/\{[p_{p\_hat}(1 - p_{p-hat})]^{1/2} *[(1/n_1) + (1/n_2)]^{1/2}\}$, wherein the proportions compared are $p_{1-hat} = (x_1/n_1)$ and $p_{2-hat} = (x_2/n_2)$, and $p_{p-hat} = (x_1 + x_2/n_1 + n_2)$. Z >1.645 is significant at $P < 0.05$ when using a one-tailed test. Z >1.960 is significant at $P < 0.05$ when using a two-tailed test.
[b]Hypotheses with only one possible direction of difference were examined with a one-tailed test. Specifically, the addition of claims data for event ascertainment by inclusion of both Medicare parts or by an expanded capture window could increase sensitivity or produce no change but not reduce sensitivity; the requirement for additional confirmatory diagnosis claims could reduce or not alter sensitivity achieved with single claims but not increase sensitivity.
[c]The difference in sensitivity between procedures and diagnoses was examined with a more conservative two-tailed test.

climate, alternative measures such as billing claims will likely assume growing importance as measures of clinical events in epidemiologic research among kidney transplant recipients. To date, there are limited data on the accuracy of claims-based measures in this population. We investigated the ability of Medicare billing claims algorithms to detect cardiovascular events recorded in the electronic medical records of one transplant program and observed several key findings: (1) Medicare Part A and Part B claims showed similar but nonoverlapping performance, such that overall sensitivity for diagnoses and procedures was statistically superior with use of algorithms combining claims from both institutions and physician/suppliers; thus, components of unique clinical information are captured in the two billing sources; (2) Perfect capture of the four studied procedures was possible within ±30 d or with short claims window expansion; in contrast, maximal sensitivity for all diagnoses was somewhat less than 100% with all study algorithms; (3) expansion of claims capture window up to ±90 d did not increase sensitivity for the collection of cardiovascular events overall but produced a trend toward improved capture for two potentially chronic diagnoses; and (4) the requirement for additional confirmatory claims did not appreciably reduce sensitivity of claims for the diagnoses of interest.

It is notable that perfect capture of clinical events by billing claims in this study was possible only for the study procedures. This finding is logically consistent with the primary purpose of claims codes to support reimbursement. Claims may be most sensitive for events that maximize reimbursement, including expensive surgeries and procedures, but be more prone to miss some diagnoses. Although all of the cardiovascular diagnoses studied are clinically important, claims seem unlikely to capture completely all diagnoses in complex patients such as transplant recipients, who often have multiple active diagnoses present at a single care encounter. Notably, the first two diagnosis codes for ambulatory visits in our transplant clinic are generally allocated to indicate management of the renal allograft (V42.0) and immunosuppression (V58.69). The lower sensitivity of diagnoses may also be related to the intermittency of visits for chronic conditions, whereas procedures are often managed as discrete events. Both the study of Hebert *et al.* (9) and a more recent investigation of national Veterans Affairs administrative data (27) found that a 2-yr surveillance period maximized the sensitivity of claims for diagnosis of diabetes.

Ideally, dates on billing claims should correspond closely in time to dates of clinically recorded events. The observed need for expanded ascertainment windows to maximize capture of atrial fibrillation and venous thromboembolism in this analysis likely reflects both the nature of the clinical reference and the types of events. Date entry for some events in the clinical record was performed as month and year, so a minimum ±30 d capture window was chosen as a primary study definition. Furthermore, retrospective ascertainment of events may have produced some imprecision in the clinical reference. We previously found that expanding the detection window in Medicare claims from ±30 d to ±90 d improved concordance of pharmacy claims with both electronic medical records and OPTN survey reports (28). Notably, widening the claims capture win-

dow had modest impact on the sensitivity of claims for the overall collection of cardiovascular procedures and diagnoses in this study, producing no change for two thirds of the events. The number of captured atrial fibrillation and venous thromboembolism events, potentially chronic cardiovascular conditions, was larger with the expanded window, although statistical significance of the change was limited by the small sample of reference events. Aside from imprecision in date recording, it is possible that events that may be managed in multiple clinical encounters were present at discrepant dates recorded in the medical records and claims but reached priority for clinical recording and billing at somewhat different times; however, this subanalysis was not a primary hypothesis and requires examination with a larger sample.

Implications of the observed deficits in the sensitivity of claims for cardiovascular diagnoses can be framed by considering claims-based estimates of disease frequency from previously published epidemiologic studies. For example, recent studies using single Part A or B Medicare claims as event measures among national samples of kidney transplant recipients reported 1-yr cumulative incidences of 3.0% for stroke/transient ischemic attack (21) and 5.6% for myocardial infarction (19), whereas the 1-yr incidence of new-onset congestive heart failure was 10.2% by the Hebert method (15). Assuming generalization of the sensitivity estimates from this study, the corrected incidences of these events would be 3.4, 6.7, and 11.0%, which do not represent marked changes for population-based estimates. Furthermore, a substantial focus of claims-based epidemiology is to estimate relative risks for events in relation to baseline factors of interest, and relative measures would not be affected by submaximal sensitivity unless it is differential with respect to a factor of interest.

An ideal surrogate measure optimizes both sensitivity and specificity for reference events, but often there is a tradeoff between sensitivity and specificity, and the relative importance of these metrics of measurement accuracy may depend on intended use. Given the nature of the available reference source of cardiovascular events, in which some cardiovascular events that were managed in patients' home communities may have escaped recording in the center's database, we studied the ability of claims to detect documented events as one dimension of measure performance in this analysis. A previous comparison of single-hospital claims for myocardial infarction with hospital charts among general Medicare beneficiaries documented high positive predictive value of claims but did not report other metrics of measurement accuracy such as sensitivity (22). Although we did not examine specificity in this analysis, the requirement for additional confirmatory diagnosis claims as in the Hebert method is an approach designed to improve specificity by reducing capture of cases in which a diagnosis is under consideration but not yet confirmed (9). Notably, application of a more stringent, adapted-Hebert case definition did not appreciably reduce the sensitivity of single claims for cardiovascular diagnoses in our analysis. In a previous comparison of billing claims in Canada with self-reported diagnoses of cardiovascular health measures including myocardial infarction, "other heart disease," and stroke, the require-

ment for more than one claim was universally associated with decrements in measure concordance (a measure of total agreement adjusted for chance) (29).

Our study is limited by availability of only Medicare claims for comparison with the clinical database. Medicare claims are particularly relevant to research among kidney transplant recipients because, unlike the eligibility requirements of age ≥65 or disability in the general population, renal allograft recipients are offered disease-specific Medicare entitlement, and Medicare is the largest single insurer in this population; however, our findings may not generalize to the administrative claims of other insurance systems. Comparisons based on small samples face risk for type 2 errors, but our sample size was adequate for detection of some important differences. Furthermore, similarities in the magnitudes of many proportions that were statistically similar in this analysis suggest that differences that could prove statistically significant in larger samples may not be meaningful in practical terms.

Our analyses used clinical records from one center as reference measures. Extension of our analyses using aggregated clinical data from multiple centers is warranted to assess the robustness of these findings. The reason for limited research on the accuracy of billing claims as study measures in dialysis and transplant populations is not the importance of the topic but rather likely reflects the technical complexity of assembling suitable data. Most payers, including Medicare, are not care providers and maintain databases limited to records of claims and benefits eligibility. Validation of claims-based research measures requires linkage of claims data to external data containing alternative measures of specific diagnoses and procedures of interest, such as electronic medical records, clinical databases, or surveys. Along with the need for electronically recorded alternative measures, requirements for access to identifying individual information to implement unique patient-level data links pose challenges for measures agreement studies with administrative data. Our study of claims sensitivity using linked single-center clinical records and Medicare claims is an important step that we hope will inspire multicenter, collaborative assessments of the accuracy of claims as research measures in transplantation.

Conclusions

This investigation of a novel linkage of Medicare billing claims and clinical records demonstrates generally good sensitivity of claims for cardiovascular events among kidney transplant recipients. Our findings support use of combined Part A and B claims to maximize sensitivity and suggest that the sensitivity of claims may be somewhat higher for procedures than for diagnoses. Administrative data provide measures for conducting timely, cost-effective, and unobtrusive research on large populations. Future collaborative and ideally prospective study is warranted to define algorithms that maximize specificity as well as sensitivity of claims from Medicare and other insurers as research measures in this population.

Disclosures
None.

References

1. Kasiske BL, Guijarro C, Massy ZA, Wiederkehr MR, Ma JZ: Cardiovascular disease after renal transplantation. *J Am Soc Nephrol* 7: 158–165, 1996
2. US Renal Data System: *USRDS 2008 Annual Data Report*, Bethesda, National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, 2008
3. Fewer Required Elements on OPTN/UNOS Forms. OPTN News Release, February 7, 2007. Available at: http://www.unos.org/SharedContentDocuments/Data_Reduction_Flier.pdf. Accessed January 5, 2009.
4. Cosimi AB, Fine RN: Proposed modification to data elements on UNetSM Transplant Candidate Registration (TCR), Transplant Recipient Registration (TRR), Transplant Recipient Follow-up (TRF), and Post-Transplant Malignancy (PTM) forms by the Policy Oversight Committee, June 12, 2006. Available at: http://asts.org//Advocacy/Regulatory01.aspx. Accessed January 5, 2009.
5. Centers for Medicare and Medicaid Services, Coding. Available at: http://www.cms.hhs.gov/home/medicare.asp. Accessed January 5, 2009.
6. Jencks SF, Cuerdon T, Burwen DR, Fleming B, Houck PM, Kussmaul AE, Nilasena DS, Ordin DL, Arday DR: Quality of medical care delivered to Medicare beneficiaries: A profile at state and national levels. *JAMA* 284: 1670–1676, 2000
7. Schneider EC, Zaslavsky AM, Epstein AM: Quality of care in for-profit and not-for-profit health plans enrolling Medicare beneficiaries. *Am J Med* 118: 1392–1400, 2005
8. Betancourt JA, Hakre S, Polyak CS, Pavlin JA: Evaluation of ICD-9 codes for syndromic surveillance in the electronic surveillance system for the early notification of community-based epidemics. *Mil Med* 172: 346–352, 2007
9. Hebert PL, Geiss LS, Tierney EF, Engelgau MM, Yawn BP, McBean AM: Identifying persons with diabetes using Medicare claims data. *Am J Med Qual* 14: 270–277, 1999
10. Kasiske BL, Snyder JJ, Gilbertson D, Matas AJ: Diabetes mellitus after kidney transplantation in the United States. *Am J Transplant* 3: 178–185, 2003
11. Kasiske BL, Snyder JJ, Gilbertson DT, Wang C: Cancer after kidney transplantation in the United States. *Am J Transplant* 4: 905–913, 2004
12. Abbott KC, Hypolite IO, Hshieh P, Cruess D, Taylor AJ, Agodoa LY: Hospitalized congestive heart failure after renal transplantation in the United States. *Ann Epidemiol* 12: 115–122, 2002
13. Abbott KC, Lentine KL, Bucci JR, Agodoa LY, Koff JM, Holtzmuller KC, Schnitzler MA: Impact of diabetes and hepatitis after kidney transplantation on patients who are

affected by hepatitis C virus. *J Am Soc Nephrol* 15: 3166–3174, 2004

14. Abbott KC, Swanson SJ, Richter ER, Bohen EM, Agodoa LY, Peters TG, Barbour G, Lipnick R, Cruess DF: Late urinary tract infection after renal transplantation in the United States. *Am J Kidney Dis* 44: 353–362, 2004

15. Lentine KL, Schnitzler MA, Abbott KC, Li L, Burroughs TE, Irish W, Brennan DC: *De novo* congestive heart failure after kidney transplantation: A common condition with poor prognostic implications. *Am J Kidney Dis* 46: 720–733, 2005

16. Lentine KL, Schnitzler MA, Abbott KC, Li L, Xiao H, Burroughs TE, Takemoto SK, Willoughby LM, Gavard JA, Brennan DC: Incidence, predictors, and associated outcomes of atrial fibrillation after kidney transplantation. *Clin J Am Soc Nephrol* 1: 288–296, 2006

17. Snyder JJ, Kasiske BL, Maclean R: Peripheral arterial disease and renal transplantation. *J Am Soc Nephrol* 17: 2056–2068, 2006

18. Dobbels F, Skeans MA, Snyder JJ, Tuomari AV, Maclean JR, Kasiske BL: Depressive disorder in renal transplantation: An analysis of Medicare claims. *Am J Kidney Dis* 51: 819–828, 2008

19. Lentine KL, Brennan DC, Schnitzler MA: Incidence and predictors of myocardial infarction after kidney transplantation. *J Am Soc Nephrol* 16: 496–506, 2005

20. Kasiske BL, Maclean JR, Snyder JJ: Acute myocardial infarction and kidney transplantation. *J Am Soc Nephrol* 17: 900–907, 2006

21. Lentine KL, Rey LA, Kolli S, Bacchi G, Schnitzler MA, Abbott KC, Xiao H, Brennan DC: Variations in the risk for cerebrovascular events after kidney transplant compared with experience on the waiting list and after graft failure. *Clin J Am Soc Nephrol* 3: 1090–1101, 2008

22. Kiyota Y, Schneeweiss S, Glynn RJ, Cannuscio CC, Avorn J, Solomon DH: Accuracy of Medicare claims-based diagnosis of acute myocardial infarction: estimating positive predictive value on the basis of review of hospital records. *Am Heart J* 148: 99–104, 2004

23. Researcher's Guide to the United States Renal Data System Database, 2008. http://www.usrds.org/research.htm. Accessed January 5, 2009.

24. Whiting JF, Woodward RS, Zavala EY, Cohen DS, Martin JE, Singer GG, Lowell JA, First MR, Brennan DC, Schnitzler MA: Economic cost of expanded criteria donors in cadav-

eric renal transplantation: Analysis of Medicare payments. *Transplantation* 70: 755–760, 2000

25. Kasiske BL, Vazquez MA, Harmon WE, Brown RS, Danovitch GM, Gaston RS, Roth D, Scandling JD, Singer GG: Recommendations for the outpatient surveillance of renal transplant recipients. American Society of Transplantation. *J Am Soc Nephrol* 11[Suppl 15]: S1–S86, 2000

26. Abbott KC, Cruess DF, Agodoa LY, Sawyers ES, Tveit DP: Early renal insufficiency and late venous thromboembolism after renal transplantation in the United States. *Am J Kidney Dis* 43: 120–130, 2004

27. Miller DR, Safford MM, Pogach LM: Who has diabetes? Best estimates of diabetes prevalence in the Department of Veterans Affairs based on computerized patient data. *Diabetes Care* 27[Suppl 2]: B10–B21, 2004

28. Buchanan PM, Schnitzler MA, Brennan DC, Dzebisashvili N, Willoughby LM, Axelrod D, Salvalaggio PR, Abbott KC, Burroughs TE, Lentine KL: Novel methods for tracking long-term maintenance immunosuppression regimens. *Clin J Am Soc Nephrol* 3: 117–124, 2008

29. Robinson JR, Young TK, Roos LL, Gelskey DE: Estimating the burden of disease: Comparing administrative data and self-reports. *Med Care* 35: 932–947, 1997

30. Luepker RV, Apple FS, Christenson RH, Crow RS, Fortmann SP, Goff D, Goldberg RJ, Hand MM, Jaffe AS, Julian DG, Levy D, Manolio T, Mendis S, Mensah G, Pajak A, Prineas RJ, Reddy KS, Roger VL, Rosamond WD, Shahar E, Sharrett AR, Sorlie P, Tunstall-Pedoe H: Case definitions for acute coronary heart disease in epidemiology and clinical research studies: A statement from the AHA Council on Epidemiology and Prevention; AHA Statistics Committee; World Heart Federation Council on Epidemiology and Prevention; the European Society of Cardiology Working Group on Epidemiology and Prevention; Centers for Disease Control and Prevention; and the National Heart, Lung, and Blood Institute. *Circulation* 108: 2543–2549, 2003

31. Swedberg K, Cleland J, Dargie H, Drexler H, Follath F, Komajda M, Tavazzi L, Smiseth OA, Gavazzi A, Haverich A, Hoes A, Jaarsma T, Korewicki J, Levy S, Linde C, Lopez-Sendon JL, Nieminen MS, Pierard L, Remme WJ: Guidelines for the diagnosis and treatment of chronic heart failure: Executive summary (update 2005). The Task Force for the Diagnosis and Treatment of Chronic Heart Failure of the European Society of Cardiology. *Eur Heart J* 26: 1115–1140, 2005