Identification of In-Hospital Complications from Claims Data: Is It Valid?
Author(s): Ann G. Lawthers, Ellen P. McCarthy, Roger B. Davis, Laura E. Peterson, R. Heather Palmer and Lisa I. Iezzoni
Source: *Medical Care*, Vol. 38, No. 8 (Aug., 2000), pp. 785-795
Published by: Lippincott Williams & Wilkins
Stable URL: http://www.jstor.org/stable/3766959
Accessed: 06-07-2018 14:40 UTC

MEDICAL CARE
Volume 38, Number 8, pp 785–795

Identification of In-Hospital Complications From Claims Data

Is It Valid?

ANN G. LAWTHERS, ScD,* ELLEN P. MCCARTHY, PhD,† ROGER B. DAVIS, ScD,†
LAURA E. PETERSON, BSN, SM,* R. HEATHER PALMER, MB, BCh, SM,*
AND LISA I. IEZZONI, MD, MSc†

OBJECTIVES. This study examined the validity of the Complications Screening Program (CSP) by testing whether (1) ICD-9-CM codes used to identify a complication are coded completely and accurately and (2) the CSP algorithm successfully separates conditions present on admission from those occurring in the hospital.

METHODS. We compared diagnosis and procedure codes contained in the Medicare claim with codes abstracted from an independent re-review of more than 1,200 medical records from Connecticut and California.

RESULTS. Eighty-nine percent of the surgical cases and 84% of the medical cases had their CSP trigger codes corroborated by re-review of the medical record. For 13% of the surgical cases and 58% of the medical cases, the condition represented by the code was judged to be present on admission rather than occurring in-hospital. The positive predictive value of the claim was greater than 80% for the surgical risk pool, suggesting the value of the CSP as a screening tool.

CONCLUSIONS. The CSP has validity as a screen for most surgical complications but only for 1 medical complication. The CSP does not have validity as a "stand-alone" tool to identify more than a few in-hospital surgery-related events. The addition of an indicator to the Medicare claim to capture the timing of secondary diagnoses would improve the validity of the CSP for identifying both surgical and medical in-hospital events.

Key words: quality; validity; claims data; performance measurement. (Med Care 2000; 38:785–795)

In 1992, we developed the Complications Screening Program (CSP), a computerized method that uses discharge abstracts to identify potentially preventable complications of hospital care.[1–3] The development of the CSP was motivated in part by the need to produce measures of quality care that used only computer-readable data. Using codes from all sections of the ICD-9-CM (International Classification of Diseases, 9th revision, Clinical Modification) diagnosis and procedure manual, the CSP screens the claim for the occurrence of 28 complications, such as postoperative acute myocardial infarction (AMI), wound infection, or reopening of surgical site. Because of this screening

*From the Center for Quality of Care Research and Education, Department of Health Policy and Management, Harvard School of Public Health, Boston, Massachusetts.

†From the Division of General Medicine and Primary Care, Department of Medicine, Harvard Medical School, Beth Israel Deaconess Medical Center, the Charles A. Dana Research Institute, and the Harvard-Thorndike Laboratory, Boston, Massachusetts.

This research was supported by the Agency for Healthcare Research and Quality under grant No. R01 HS09099. The views expressed are solely those of the authors.

An earlier version of this article was presented as a poster at the Association for Health Services Research conference, June 21–23, 1998, Chicago, Illinois.

Correspondence to: Dr. A.G. Lawthers, Center for Quality of Care Research and Education, Harvard School of Public Health, 677 Huntington Ave, Boston, MA 02115. E-mail: alawthers@mediaone.net

Received January 11, 1999; initial review completed March 31, 1999; accepted November 9, 1999.

function, the groups of ICD-9-CM codes that identify a complication are known as "screens."

See p 777

Each complication screen is applied to 1 or more "risk pools." The risk pools identify patients who are at risk for the complication and thus serve as the denominator in the calculation of complication rates. The CSP creates 6 risks pools: major surgery, minor surgery, invasive cardiac procedures, endoscopy, medical patients, and all patients. A patient may be assigned to more than 1 risk pool (eg, major surgery and all patients).

The CSP algorithm attempts to separate conditions that arise during the hospitalization (ie, potential complications of care) from those present on admission. The ICD-9-CM diagnosis and procedure codes from the claim and number of days from admission to principal major surgeries or procedures are used to identify patients with specific complications. We call the ICD-9-CM codes "trigger codes" because the presence of a specified code causes or triggers assignment of a case to a complication screen. The computer algorithm then tests whether specified qualifying conditions are met. For the postoperative AMI screen, for example, AMI must not be the principal diagnosis, the patient must not be in major diagnostic category 5 (cardiovascular conditions), the principal surgery must not be cardiac surgery, and surgery must have occurred on the day of admission or the second hospital day.

These qualifying statements aim to eliminate cases for which the code on the claim resulted from underlying disease rather than substandard care. For example, the timing criterion (surgery on the first or second day of hospital stay) aims to increase the probability that the code on the claim represents a complication. A patient hospitalized for several days before surgery (instead of only 1 or 2 days) may have developed an AMI that was treated before surgery or perhaps developed an AMI as a complication of the surgery. It is impossible to distinguish these 2 cases from the claims data alone. If all qualifying conditions are met, the case is flagged as a potential complication. The logic is available on request.

Others have also used discharge abstracts to target hospital complications,[2,4] but generally these methods use ICD-9-CM diagnosis codes that are explicitly labeled as "complications" (codes beginning with 996–999). Although these codes specifically delineate complications, whether they are coded completely remains questionable.

Our original study found that the CSP identified patients more likely to die in-hospital and who had longer lengths of stay and significantly higher charges than other cases.[1–3] However, given concerns about the accuracy of discharge abstracts and the timing of diagnoses, we designed a 3-part study to validate the CSP. The 3 parts would validate the data used to construct the screens, validate the screens as a flag for actual quality problems, and validate the replicability of hospital-level results using different data sources. The study we report here focuses on the validity of the data and addresses 2 questions: (1) Are the ICD-9-CM codes used to trigger a CSP complication flag coded completely and accurately? (2) Does the CSP algorithm successfully separate conditions present on admission from those that occur in the hospital? We examined validity by comparing the codes contained in the claim with codes abstracted during an independent re-review of the medical record.

Methods

Sample

We conducted the validation study with the Peer Review Organizations (PROs) in 2 states, Connecticut and California, using Medicare's fiscal year 1994 MEDPRO database. Our sample was not designed to demonstrate variation between hospitals within a state. Additional details of our sampling approach are described elsewhere.[5,6]

We limited our validation to the major surgery and medical risk pools because these encompass the majority of hospital cases. For each hospital, we estimated their expected complication rate, using logistic regression models that corrected for the effect of patient risk factors on the probability of a discharge being flagged by the CSP. The models were applied separately to the 2 risk pools, combining data from California and Connecticut. We computed z scores, using the observed and expected rates of complications, plus the standard error of the expected rate. A negative z score indicated that fewer cases were flagged than expected under the model, and a positive score indicated more than expected. We then ranked hospitals by z scores.

We next defined 3 strata of hospitals within each risk pool for each state: high, middle, and low z scores. In California, we followed our original plan of taking the top 15%, middle 30%, and bottom 15% of the hospitals in each risk pool. Because of the small number of Connecticut hospitals, we defined their 3 strata as hospitals with the 9 lowest scores, the 11 middle scores, and the 9 highest scores for major surgery and the 7 lowest scores, the 9 middle scores, and the 7 highest scores for medical cases. Within each stratum, we selected hospitals at random: 8 per stratum for general surgery and 6 per stratum for medical cases. Because the same hospital could be sampled for both the surgical as well as the medical risk pool, we sampled fewer than the theoretically possible number of hospitals in each state. We sampled a total of 28 hospitals for Connecticut and 41 hospitals for California.

Next we selected cases within hospitals. Some of the 28 original CSP screens occurred too infrequently for rigorous validation. Therefore, we validated 17 of the original 28 screens for major surgery and 6 of the original 7 screens for medical. Within each hospital selected from the major surgery risk pool, we sampled up to 20 discharges: 1 at random from among those flagged as a complication by each of 17 individual major surgery screens and 3 at random from among those that were not flagged by any screen (ie, control cases). Because some screens were not triggered in any cases at some hospitals, fewer than 20 discharges were sampled at certain hospitals. We selected 406 major surgery discharges in California and 405 in Connecticut. Similarly, we sampled up to 16 discharges from each medical risk pool hospital: 2 at random from among those flagged by each of 6 medical screens and 4 unflagged discharges (ie, control cases). A total of 236 medical risk pool discharges were selected in California and 256 in Connecticut. Appendix A includes figures showing cases sampled and reviewed for both the study reported here and subsequent analyses involving explicit and structured implicit review (summarized in separate publications).

The final sample lists were forwarded to the PROs in each state, which contacted the hospitals and obtained photocopies of the patient records. At this point, we discovered that some hospitals had closed. Additional hospitals and cases were sampled to replace the closed institutions and their cases.

Data Collection Process

To collect data, we specially designed an abstraction tool that was computerized by the California PRO. Reviews were completed in 2 stages by Accredited Record Technicians trained in the instrument's use. During stage 1, reviewers abstracted basic demographic information about the case and then recoded the medical record using standard American Hospital Association guidelines for medical record coding. Reviewers were instructed not to use information from the physician attestation or the list of diagnoses and procedures coded by the hospital's medical records department but rather to conduct an independent abstraction from the appropriate components of the medical record. The reviewer entered the principal diagnosis, principal procedure, and up to 14 additional diagnoses and 14 additional procedures for each case. For each diagnosis, the reviewer answered the question, "Was the diagnosis present on admission?" Reviewers were instructed to code "yes" if the diagnosis was documented in emergency room notes or admission notes that were dated no later than 1 day after the date of admission. The reviewer then answered the question, "If the diagnosis was not present on admission, when was the diagnosis first noted?" The reviewer also recorded the date of each procedure. When the initial abstraction was complete, the reviewer pressed a key to begin stage 2 of the review. Once stage 2 had begun, the computer program prevented the reviewer from going back and accessing or changing stage 1 data.

In stage 2, the reviewer was presented with a list of all the codes abstracted during stage 1, plus any trigger codes from the original Medicare claim (ie, codes assigned by the hospital that are used by the CSP to flag complications). The reviewer was instructed to review the trigger codes and assign 1 of 3 values to each trigger code: (1) the code is an exact match with the reviewer's stage 1 abstraction, (2) the re-review of the record reveals information that now justifies the trigger code, or (3) the trigger code is incorrect, that is, the documentation supports a different code or finds no corroborating evidence for the code.

Data quality checks included a re-review for interrater reliability of a randomly selected 15% of cases. We limited our interrater analysis to trigger codes used by the CSP and to the principal diagnosis and procedure. For CSP trigger codes, if one reviewer coded a CSP complication such as

pulmonary compromise (ICD-9-CM codes 514, 518.4, 518.5, 518.81, and 518.82), we considered agreement to be present if the other reviewer coded any of the 5 codes, rather than requiring the identical codes.

Analyses

The analyses focused on establishing validity of the Medicare claim as the data source for flagging in-hospital complications by use of the CSP screens and on the CSP's presumptions about timing. We assessed validity by (1) corroborating CSP trigger codes by medical record review, (2) checking the accuracy of the CSP timing assumptions, (3) calculating whether a claim code predicted the presence of documentation supporting the code in the medical record (predictive values of a claim), and (4) measuring interrater reliability (because a measure cannot be valid unless it is also reliable).

First we created a "gold standard" set of ICD-9-CM codes for each case by enriching the list of codes found during stage 1 review with any additional trigger codes identified during stage 2 (see above). Next, we applied the CSP to the gold standard codes to identify the presence of a complication based on the recode of the medical record. We then compared the 2 sets of CSP results: those based on medical record data versus those based only on Medicare claims data.

For those cases in which the presence of a complication was corroborated by medical record abstraction, we evaluated the timing of key events. Was the code, which signaled a complication, actually present on admission and therefore preexisting rather than an in-hospital complication? Was the index procedure performed on day 1 or day 2 of the hospital stay for surgical cases? Did the documentation of the diagnosis signaling a complication first appear after the index procedure for surgical cases (ie, postoperatively)? Finally, we calculated the overall percentage of cases for each complication screen that met the CSP assumptions (codes and timing). This final percentage estimated how often a case flagged by the CSP was found to have an in-hospital event.

We calculated the predictive values, positive and negative, of a claim. We calculated positive predictive value (PPV) as the number of CSP-flagged cases with confirmed complication divided by the number of CSP-flagged cases. Negative predictive value (NPV) was calculated as the number of unflagged cases with confirmed lack of complication divided by the number of unflagged cases. We used all the codes identified during the recode of the medical record, including codes identified among the controls, to count the number of cases with a confirmed complication. This means that a case could be assigned to multiple complication screens instead of just 1 screen, as used for other analyses. High PPV validates the CSP as a tool for identifying cases likely to have documentation supporting the claims code in the medical record. Predictive value should not be confused with the percent of cases confirmed as in-hospital events.

All analyses were conducted with SAS version 6.12.[7]

Results

Characteristics of the Study Sample

This study involved 1,298 cases, 634 (48.8%) from California and 664 (51.2%) from Connecticut. Of these, 813 (62.6%) were surgical cases and 485 (37.4%) were medical cases. Among surgical cases, the mean (SD) age was 76.2 (7.1) years, 53.3% were women, and 7.6% were nonwhite. Among medical cases, the mean (SD) age was 77.6 (7.5) years, 58.1% were women, and 14.9% were nonwhite.

Confirmation of Trigger Codes and Timing Assumptions

Corroboration of Trigger Code by Medical Record Review. Table 1 presents the results of the corroboration of trigger code by medical record reabstraction, confirmation of timing assumptions, and overall proportion of cases confirmed as in-hospital complications (ie, cases for which the code was corroborated and the timing assumptions were confirmed). The results are sorted by the overall proportion of cases confirmed. Screens with the highest proportion of cases with trigger codes corroborated on record review included surgical iatrogenic complications (98%), reopening of surgical site (97%), and mechanical complications (96%). Screens with the lowest proportion of cases with corroborated codes were miscellaneous complications (67%) and medication complications (67%).

TABLE 1. Corroboration on Re-Review of Codes and Timing Assumptions Used by the CSP, Ordered by Cases Confirmed

	Cases Reviewed: a	Cases With Codes Corroborated: b (% = b/a)	Diagnosis Present on Admission: c (% = c/b)	Other Timing Problems*: d (% = d/b)	Cases With Codes and In-Hospital Timing Confirmed: e = b − c − d (% = e/a)
Surgical risk pool					
Reopening of surgical site	32	31 (97)	0 (0)	0 (0)	31 (97)
Iatrogenic complications	48	47 (98)	0 (0)	3 (6)	44 (92)
Postoperative wound infection	44	41 (93)	0 (0)	1 (2)	40 (91)
Postoperative AMI	37	34 (92)	3 (9)	0 (0)	31 (84)
Postoperative cerebral infarction	35	29 (83)	0 (0)	0 (0)	29 (83)
Postprocedural hemorrhage or hematoma	46	42 (91)	1 (2)	5 (7)	38 (83)
Aspiration pneumonia	35	33 (94)	5 (15)	1 (3)	27 (77)
Postoperative infection (except wound and pneumonia)	32	30 (94)	7 (23)	0 (0)	23 (72)
Postoperative pulmonary compromise	46	42 (91)	5 (12)	4 (10)	33 (72)
Procedure-related perforation or laceration	35	33 (94)	8 (24)	0 (0)	25 (71)
Postoperative gastrointestinal hemorrhage	41	33 (81)	5 (15)	1 (3)	27 (66)
Postoperative mechanical complications	46	44 (96)	13 (30)	3 (7)	28 (61)
Deep vein thrombosis/pulmonary embolism	41	36 (88)	8 (22)	4 (11)	24 (59)
In-hospital hip fracture and falls	21	19 (91)	4 (21)	3 (16)	12 (57)
Postoperative shock or cardiorespiratory arrest	40	34 (85)	10 (29)	3 (9)	21 (53)
Miscellaneous complications	46	31 (67)	6 (19)	2 (6)	23 (50)
Total surgical risk pool	625	559 (89)	75 (13)	28 (5)	456 (73)
Medical risk pool					
Iatrogenic complications	68	55 (81)	15 (27)	0 (0)	40 (59)
Postprocedural hemorrhage or hematoma	53	48 (91)	15 (31)	7 (15)	26 (49)
Deep vein thrombosis/pulmonary embolism	68	53 (78)	30 (57)	1 (2)	22 (32)
Wound infection	39	37 (95)	25 (68)	1 (3)	11 (28)
In-hospital hip fracture or fall	64	62 (97)	54 (87)	1 (2)	7 (11)
Medication complications	51	34 (67)	28 (82)	1 (3)	5 (10)
Total medical risk pool	343	289 (84)	167 (58)	11 (4)	111 (32)

Values are n or n (%).
*Surgery after day 2, complication before surgery or procedure, or unable to confirm complication date.

Confirmation of Timing Assumptions. Columns c and d of Table 1 show the results of the verification of the timing assumptions for the surgical and medical cases. Overall, we found that the CSP trigger code was present on admission for 13% of the surgical cases and 58% of the medical

cases. This means that the trigger code represented a preexisting condition rather than one that arose during hospitalization. For example, for surgical cases, 30% of the mechanical complications were present on admission, as was shock in 29% of the cases. Among medical cases, 87% of hip fractures or falls were present on admission, as were 82% of the medication complications.

The difference between risk pools for the same complication is also striking. For example, nearly 68% of wound infections in medical cases were present on admission, whereas none of the wound infections in the surgical cases were present on admission.

Overall Confirmation of Complications. The overall proportion of cases confirmed as in-hospital events (Table 1, last column) ranged from a low of 10% (medication complications, medical cases) to a high of 97% (reopening of surgical site, surgical cases). The proportion of complications confirmed for surgical screens was considerably higher than for the medical screens (73% vs. 32%).

The proportion of cases confirmed differed somewhat by state. Thirty percent of the medical cases were confirmed for Connecticut, whereas only 15% of the California medical cases were confirmed (Fisher's exact test, $P < 0.0001$). However, for the surgical risk pool, the proportion of cases confirmed did not differ statistically by state (70% for Connecticut and 76% for California; $P = 0.09$).

Predicting the Presence of Documentation Supporting a Complication Code

Table 2 presents the predictive values of each CSP screen. A PPV answers the question, "If the CSP flags a discharge for a specific complication based on the claim, what is the probability that a review of the medical record supports the coding of any complication?" An NPV answers the converse, "If a claim does not cause the CSP to flag the discharge for a specific complication, what is the probability that a review of the medical record also fails to support a complication code?"

PPV varied by screen and by risk pool. For example, for surgical cases, PPV was highest for postoperative infection (97%) and iatrogenic complications (93%) and lowest for miscellaneous complications (69%). For medical cases, iatrogenic complications had only a 75% PPV. PPVs for wound infection, deep vein thrombosis, and in-hospital hip fracture also varied across risk pools.

Interrater Reliability

Overall, percent agreement between PRO reviewers was excellent for the presence of a trigger code flagging a case into a specified screen (>90%), with the exception of the surgical screens of miscellaneous and iatrogenic complications, for which the percent agreement was 71% and 74%, respectively. On the other hand, kappas ranged widely from 0.02 for miscellaneous complications to 1.00 for postoperative cerebral infarction, with a median of 0.66 (postoperative mechanical complications). There were virtually no differences in kappas between states (0.66 for Connecticut and 0.61 for California). Agreement for principal diagnosis and procedure ranged from 48% on principal procedure in the medical cases to 72% for principal procedure in surgical cases.

Discussion

The question of validity is one of "valid for what purpose?" For the purpose of identifying in-hospital events, the surgical screens validate much better than the medical screens. The CSP technique to isolate in-hospital events (ie, requiring principal procedures to have occurred on the first or second hospital day) seemed to work for surgical cases. Six surgical screens validate particularly well in terms of code corroboration and timing assumptions, with overall confirmation rates >80%. The medical screens, on the other hand, do not validate for the purpose of identifying in-hospital events, primarily because the CSP trigger codes found in the claims data often represented a condition present on admission rather than one that arose during hospitalization.

For the purpose of screening claims for possible complications, the CSP performs much better. Screens aim to identify cases for in-depth review, not to make absolute judgments. Ten surgical screens had PPVs of ~88% or higher and would be good-to-excellent candidates as screens for complications. The only surgical screen with a substandard PPV was miscellaneous complications. The remaining surgical screens might have some utility in screening for in-hospital complications.

Only 1 medical screen appeared useful as a screening tool: postprocedural hemorrhage or hematoma. On the basis of these findings, we could not recommend using the other medical screens to screen for complications.

TABLE 2. Predictive Value of a CSP Screen, Ordered by PPV

	PPV, %	NPV, %	% Agreement Claim vs Record	Kappa for Claim vs Record
Major surgery risk pool				
Postoperative infection (except wound and pneumonia)	96.8	98.3	98.2	0.80
Iatrogenic complications	93.3	67.3	75.6	0.52
Postoperative AMI	93.0	99.1	98.7	0.88
Postoperative pulmonary compromise	92.5	96.2	95.8	0.82
Postoperative wound infection	91.5	95.0	94.8	0.69
Postprocedural hemorrhage or hematoma	89.7	93.6	93.2	0.69
Deep vein thrombosis/pulmonary embolism	89.6	98.1	97.6	0.81
Postoperative shock or cardiorespiratory arrest	89.3	97.8	97.2	0.81
Postoperative mechanical complications	89.1	96.9	96.3	0.78
Reopening of surgical site	88.2	98.0	97.6	0.75
Aspiration pneumonia	85.7	97.4	96.8	0.72
In-hospital hip fracture and falls	85.0	99.2	98.9	0.78
Postoperative cerebral infarction	84.2	99.3	98.6	0.85
Procedure-related perforation or lacerations	81.6	99.1	98.2	0.81
Postoperative gastrointestinal hemorrhage	81.4	99.1	98.1	0.81
Miscellaneous complications	69.2	86.3	84.3	0.42
Medical risk pool				
Postprocedural hemorrhage or hematoma	90.6	98.6	97.7	0.88
Wound infection	80.0	99.3	97.7	0.84
Deep vein thrombosis/pulmonary embolism	75.7	98.5	95.1	0.79
Iatrogenic complications	74.6	95.9	92.8	0.71
In-hospital hip fracture and falls	60.6	99.5	93.6	0.69
Medication complications	50.0	98.8	94.2	0.62

The Art of Coding

Our findings illustrate that coding from the medical record is an art, not a science. For example, although 85% of the cases had their codes located during stage 1 (blinded) review, for fully 15% of the cases, the trigger code was located only after the reviewer was told which trigger codes were listed on the original claim. The moderate interrater agreement scores also suggest that coding has a subjective component.

Some screens were based on codes that were easier to find than others. For example, the AMI codes were usually located the first time through the record (89% located at stage 1), had excellent interrater reliability ($\kappa = 0.824$), and had excellent agreement between the claim and the record ($\kappa = 0.882$). Other screens for which the codes were relatively easy to find include postoperative cerebral infarction, postoperative infection (excluding wound), hemorrhage and hematoma (medical cases only), and deep vein thrombosis and pulmonary embolism (medical cases only).

At the opposite end of the spectrum are the trigger codes for miscellaneous complications, a heterogeneous screen comprised of codes for air embolism, postoperative fistula, and other specified adverse events, among others. These codes were difficult to locate (44% located at stage 1 review), had poor interrater reliability ($\kappa = 0.02$), and had a PPV of only 69%. Other screens for which the codes were more difficult to locate included hip fracture (medical cases only), iatrogenic complications (surgical only), and gastrointestinal hemorrhage.

Present on Admission or Occurring During Hospitalization

Our most striking finding was the large proportion of cases for which the trigger code supported by the medical record documentation represented a diagnosis present on admission instead of a condition that arose during hospitalization. To be flagged under CSP rules, the trigger code must be listed as a secondary diagnosis, that is, a condition deemed, after study, not to have been the principal reason for the admission. ICD-9-CM coding guidelines for Medicare require the coding of "all diagnoses that affect the current hospital stay," including the principal diagnosis and "other diagnoses" as follows: "Other diagnoses are designed and defined as all conditions that coexist at the time of admission, that develop subsequently, or that affect the treatment received and/or length of stay. Diagnoses that relate to an earlier episode that have no bearing on the current hospital stay are excluded."[8]

Fifty-eight percent of the medical cases had the trigger diagnosis present on admission. Far fewer surgical cases had the trigger code present on admission.

The problem of determining the timing of condition onset could be addressed fairly simply by the addition of a modifier digit to secondary diagnosis codes, as proposed by the National Committee on Vital and Health Statistics[9] and as already in place in California and New York. The data element would flag whether a diagnosis preceded or followed admission to the hospital. This simple step could dramatically improve the confirmation rates reported here and generally improve the usefulness of discharge abstracts for assessing quality of care.

Limitations of the Study

For individual screens, the number of cases examined was relatively small, although it was sufficient to make reasonable assessments about validity. Because of financial constraints, we could not examine all codes used by the CSP algorithm, such as the codes used to define the qualifying conditions or the chronic conditions used in the logistic regression models. In stage II of the reviews, we "unblinded" reviewers only to the trigger codes, admittedly the most crucial aspect of the CSP logic.

Although this study confirmed timing and events from the perspective of ICD-9-CM coding rules, it neither looked for clinical evidence that the event had actually happened (eg, for AMI, was confirmatory evidence present using symptoms, laboratory, and diagnostic test data?) nor examined whether substandard quality of care was associated with identified complications. These topics are the subjects of other papers.

One final caveat: one might argue that the data in Table 1 should be weighted to account for our sampling scheme. However, because of the complexities of the second stage of the sampling process (sampling first from the least frequently flagged screen, then the second least frequently flagged screen, then the third rarest, and so on), the calculation of appropriate weights was essentially intractable. Furthermore, it is unlikely that weighting would change our conclusions.

Summary

The Minnesota law restricting access to medical records for research and quality-assurance purposes by expanding provisions for patient notification and consent[10] may be a harbinger of the future. If it is, such requirements almost certainly will have a chilling effect on quality measurement and improvement efforts. Therefore, it is essential to find valid tools to assess the quality of care in a nonintrusive manner.

The CSP, used appropriately as a screen, can assist in flagging in-hospital events that might be complications of care. The identified cases should then be followed up by in-depth review to confirm the complication. For this purpose, surgical screens had greater validity than medical screens.

Use of the CSP to definitively identify in-hospital complications with only claims data is problematic. The difficulty is determining whether a code represents a preexisting condition present on admission or a condition arising from the hospitalization. For this purpose, only selected surgical screens were validated. The addition of an indicator of the timing of secondary diagnosis would significantly increase the utility of claims data for identifying in-hospital complications.

Acknowledgments

The authors wish to thank Kathleen Warren of the Connecticut Peer Review Organization and staff at the

California Medical Review, Inc (CMRI) who coordinated the medical record reabstraction review. Additional thanks to the CMRI technical and programming staff for their role in programming the computerized abstraction tool. The authors wish to thank Donald T. Davies, Jr, BA, and Saul N. Weingart, MD, PhD, for comments on earlier drafts of the paper.

References

1. **Iezzoni LI, Daley J, Foley SM, Heeren T, Hughes JS, Fisher ES, et al.** Screening quality of care using administrative data: Final report. Prepared for the Agency for Health Care Policy and Research under grant No. RO1-HS06512. Boston, Mass: Beth Israel Hospital, December 1992:22.

2. **Iezzoni LI, Daley J, Heeren T, Foley SM, Fisher ES, Duncan C, et al.** Identifying complications of care using administrative data. Med Care 1994;32:700–715.

3. **Iezzoni LI, Daley J, Heeren T, Foley SM, Hughes JS, Fisher ES, et al.** Using administrative data to screen hospitals for high complication rates. Inquiry 1994;31:40–55.

4. **Kalish RL, Daley J, Duncan CC, Davis RS, Coffman GA, Iezzoni LI.** Costs of potential complications of care for major surgery patients. Am J Med Qual 1995;10:48–54.

5. **Iezzoni LI, Davis RB, Palmer RH, Cahalane M, Hamel MB, Mukamal K, et al.** Does the Complications Screening Program flag cases with process of care problems? Using explicit criteria to judge processes. Int J Qual Health Care 1999;11:107–118.

6. **Iezzoni LI, Lawthers AG, Davis RB, Palmer RH, Cahalane M, Mukamel K, et al.** Screening quality of care using administrative data: Final report. Prepared for the Agency for Health Care Policy and Research under grant No. RO1-HS09099. Boston, Mass: Beth Israel Deaconess Medical Center, January 1999:31.

7. The SAS system for Windows. Version 6.12, Cary, NC: The SAS Institute, 1994, 1995.

8. **Brown F.** ICD-9-CM coding handbook, without answers. Chicago, Ill: American Hospital Publishing Inc, Division of Quality Control Management of the American Hospital Association, 1989:22.

9. National Committee on Vital and Health Statistics, Subcommittee on Ambulatory and Hospital Care Statistics. Proposed revision to the uniform hospital discharge data set. Washington, DC: National Committee on Vital and Health Statistics, 1992.

10. **Melton LJ III.** The threat to medical-records research. N Engl J Med 1997;337:1466–1470.

Appendix A. Cases Sampled and Reviewed, CSP Validation Studies.

Appendix A is presented as Figures 1 and 2.

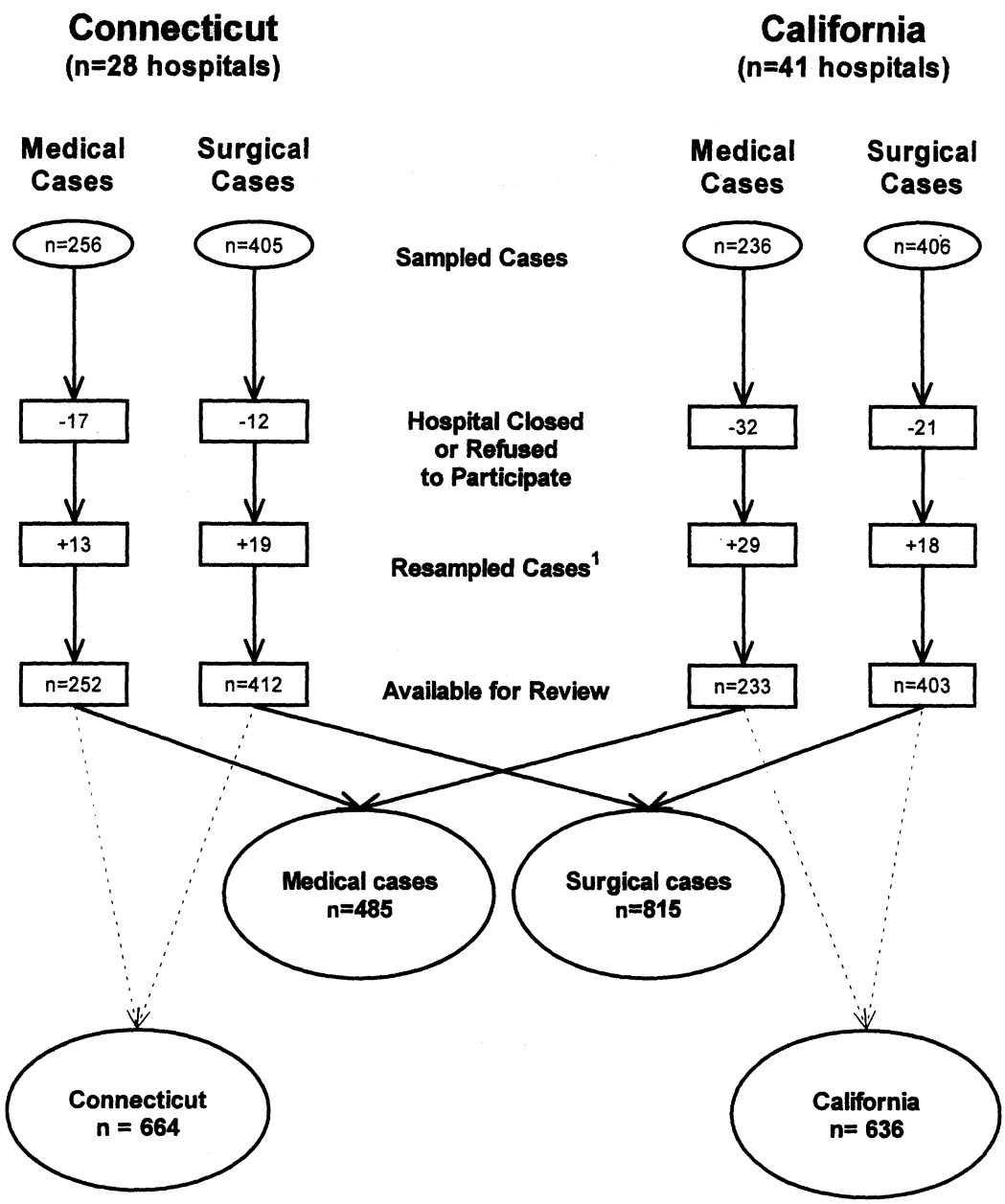

FIG. 1. Cases sampled. 1 = Resampled cases: cases sampled to replace those in hospitals that had closed or refused to participate.

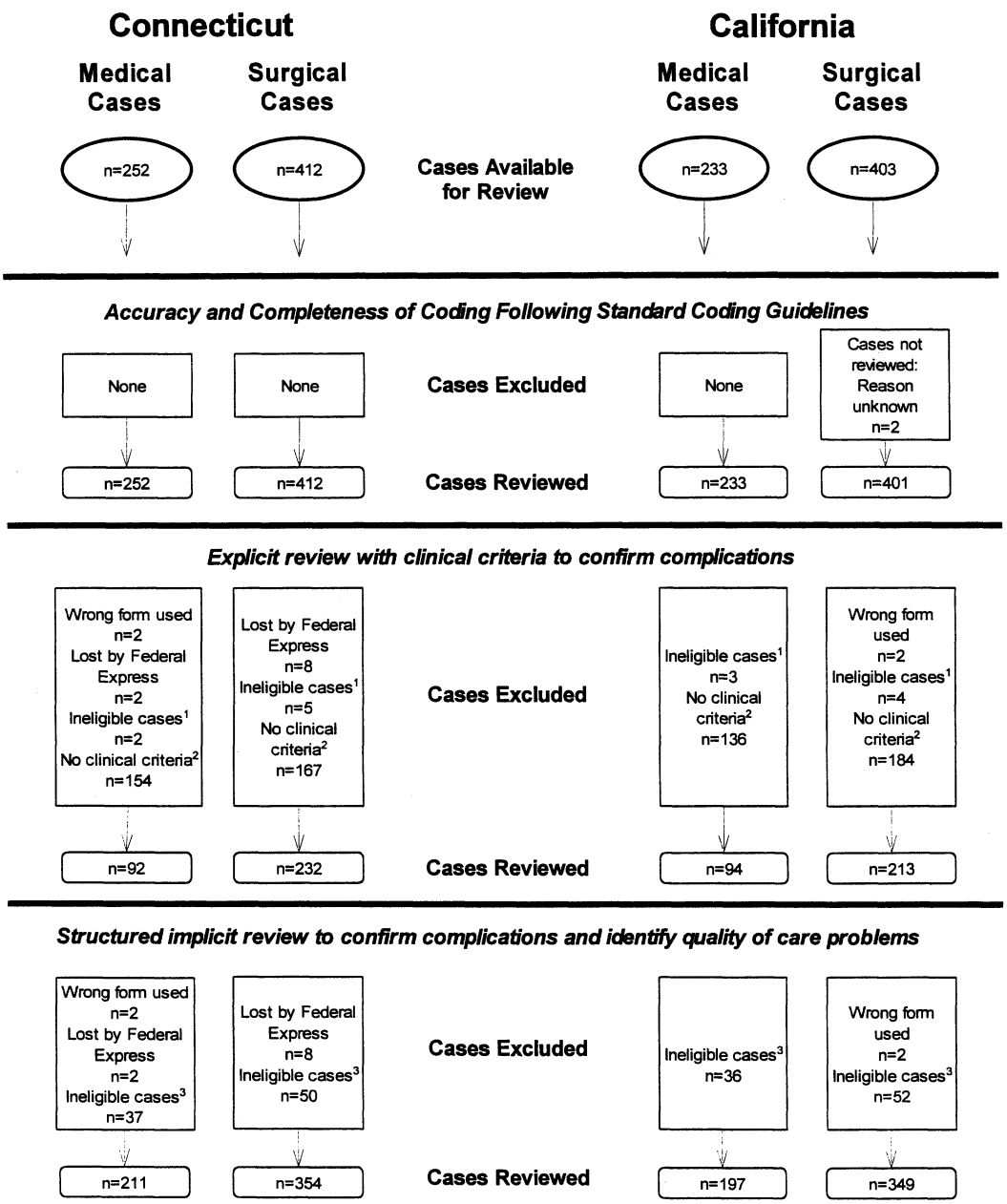

FIG. 2. Cases reviewed. 1 = Ineligible cases (cases in which the review form instructs the reviewer to stop). 2 = No clinical criteria. For some of the screens (for example, miscellaneous complications, iatrogenic complications, and medication events), we determined that creating detailed and comprehensive clinical definition criteria was not possible given our time frames. For these screens, only the processes of care were reviewed, using a "Common Processes of Care" form. 3 = Ineligible cases. These were cases for which the complication belonged to the "miscellaneous" or "iatrogenic" complication screen. The diversity of clinical situations captured by the latter 2 screens made it unfeasible to conduct an implicit review.