CHAPTER FIFTEEN: VALIDATION STUDY OF ACUTE MYOCARDIAL INFARCTION, CONCLUSIONS

This chapter briefly summarizes the key findings of the AMI validation study for each research question. Where possible, the results of the validation study are compared with previous research. Changes suggested by the validation study for future AMI mortality models of the California Hospital Outcomes Project are also described.

QUESTION 1: What proportion of cases included in the 1993 AMI study should have been excluded because acute myocardial infarction was incorrectly reported or incorrectly diagnosed?

- A total of 31 cases from the original sample of 1,005 (3.1%) were definitely false positives using the inclusion and exclusion criteria from OSHPD's 1993 report. Specifically, 18 cases had a reported principal diagnosis and 4 cases had a reported secondary diagnosis of AMI without any documentation of this diagnosis by a physician, 4 cases were post-transfer hospitalizations, and 4 cases were actually postoperative AMIs.

- Of the remaining 974 cases, 74 (7.6%) had documentation of an AMI by the treating physician but did not meet strict criteria for this diagnosis based on a history of chest pain, cardiac enzyme values, and electrocardiographic findings.

- Reweighting these figures based on the statewide population, about 2.2% of the cases included in OSHPD's 1993 AMI mortality study were definitely false positives, and an additional 7.2% were suspected to be false positives.

These estimates are substantially lower than the comparable estimates of 26% reported from 15 hospitals in the Boston area,[1] 39% reported from a major medical center in Texas,[2] and 21% reported from a Toronto teaching hospital.[3] These differences may be attributable to: (1) better physician documentation in medical records, (2) more attention to diagnostic coding since the advent of prospective payment based on Diagnosis Related Groups, (3) the introduction of a fifth digit on the ICD-9-CM code for AMI to distinguish initial from subsequent episodes of care, and (4) the use of special criteria,

---

[1] Iezzoni LI, Burnside S, Sickles L, Moskowitz MA, Sawitz E, Levine PA. Coding of acute myocardial infarction: Clinical and policy implications. *Ann Intern Med* 1988; 109:745-751.

[2] Kennedy GT, Stern MP, Crawford MH. Miscoding of hospital discharges as acute myocardial infarction: Implications for surveillance programs aimed at elucidating trends in coronary artery disease. *Am J Cardiol* 1984; 53:1000-1002.

[3] van Walraven C, Wang B, Ugnat A, Naylor CD. False-positive coding for acute myocardial infarction on hospital discharge records: Chart audit results from a tertiary center. *Can J Cardiol* 1990; 6(9):383-386.

such as a length of stay less than 4 days (3 days in the current study), to exclude patients who actually ruled out for AMI. One recent study that focused on the validity of Medicare DRGs found an 8% false positive rate.[4]

Note that this validation study did not address the number of true AMIs that were missed because of underreporting. Recent studies suggest that 76%[5] to 90%[4] of all AMIs, including postoperative and in-hospital AMIs, can be identified using ICD-9-CM diagnosis codes. This percentage is probably higher for the subset of cases admitted principally because of an AMI.

These results were used to modify the list of acceptable principal diagnoses (Table 3.1). Most notably, arterial embolism or thrombosis was removed from the 1993 list because four of the five cases with this principal diagnosis in the validation sample actually had a postoperative AMI. In OSHPD's third study of AMI mortality, now underway, complete atrioventricular block will also be removed from the list. These changes are expected to further reduce the already low false positive rate.

QUESTION 2: What is the statewide reporting accuracy for important risk factors included in the risk-adjustment models?

- The validity and reliability of coding were excellent (sensitivity>80% and κ>0.8) for infarct site and diabetes, although about 60% of patients reported to have "other or unspecified" site actually had documentation suggesting a specific site.

- The validity and reliability of coding were very good (sensitivity>60% and κ>0.6) for congestive heart failure (CHF), chronic renal disease, prior coronary bypass surgery, history of pacemaker, complete atrioventricular block, and shock.

- Several other risk factors, including epilepsy, other cerebrovascular disease, primary or secondary malignancy, and hypertension had intermediate validity and reliability ($0.45<\kappa<0.6$).

---

[4] Fisher ES, Whaley FS, Krushat WM, Malenka DJ, Fleming C, Baron JA, Hsia DC. The accuracy of Medicare's hospital claims data: Progress has been made, but problems remain. *Am J Public Health* 1992; 82:243-248.

[5] Jollis JG, Ancukiewicz M, DeLong ER, Pryor DB, Muhlbaier LH, Mark DB. Discordance of databases designed for claims payment versus clinical information systems: Implications for outcomes research. *Ann Intern Med* 1993; 119:844-850.

- Six risk factors (chronic liver disease, hypotension, late effects of cerebrovascular disease, pulmonary edema, nutritional deficiency, and other valve disease) were poorly coded (sensitivity<40% and κ<0.45).

These numbers are generally consistent with those reported elsewhere. For example, Jollis et al at Duke University Medical Center reported a sensitivity of 83% with a kappa of 0.83 for diabetes, a sensitivity of 65% with a kappa of 0.56 for hypertension, a sensitivity of 44% with a kappa of 0.48 for mitral insufficiency, a sensitivity of 36% with a kappa of 0.39 for CHF, and a sensitivity of 14% with a kappa of 0.19 for cerebrovascular disease.[5] Based on 1985 Medicare data, Fisher et al reported sensitivities of 84% for diabetes, 82% for hypertension, 89% for CHF, 92% for cerebrovascular disease, and 83% for chronic renal failure.[4] In a previous, unblinded reabstraction study of California discharge abstracts, the sensitivity of coding was 88% for diabetes (κ=0.89), 65% for hypertension (κ=0.77), 88% for chronic renal failure (κ=0.85), and 100% (κ=1.00) for chronic liver disease.[6]

These results will be used to modify the list of risk factors for OSHPD's third analysis of AMI mortality, now underway. The six risk factors described as "poorly coded" will no longer be used in risk-adjustment models. In the meantime, OSHPD will continue its intensive educational efforts designed to improve reporting of all diagnoses that affect inpatient treatment.

QUESTION 3: Are important risk factors coded more thoroughly at hospitals with low risk-adjusted mortality than at hospitals with high risk-adjusted mortality? If so, does the variation in risk-adjusted mortality diminish when inter-hospital differences in risk factor coding are removed?

- There were no consistent differences in the coding of specific risk factors across hospital mortality and volume categories, although some variation exists.

- Overall, 65.0% of the original discharge abstracts had at least one missing clinical risk factor and 30.9% had at least two missing risk factors. This percentage did not differ across hospital mortality categories, but was higher at high-volume hospitals than at medium-volume hospitals (68.8% versus 61.2%).

- Conversely, 31.5% of the original discharge abstracts had at least one unsupported clinical risk factor based on CMRI's reabstraction. This finding was more frequent at low-mortality hospitals than at intermediate or high-mortality hospitals (36.7% versus 29.2% and 29.0%, respectively), but was unrelated to hospital volume.

- Using Model B, the difference in risk-adjusted mortality between low-mortality and high-mortality outlier hospitals shrinks by 19% to 29% (on the derivation of the

[6] Romano PS, Mark DH. Bias in the coding of hospital discharge data and its implications for quality assessment. *Med Care* 1994; 32:81-90.

regression coefficients) when the original data are replaced by reabstracted CMRI data. Using Model A, the same difference shrinks by only 0% to 12%.

- Hospitals designated in the 1993 report as having low risk-adjusted mortality still have low mortality even after adjusting for all of the additional risk factors discovered through reabstraction. Although hospitals with high risk-adjusted mortality no longer appear significantly worse than expected, this finding is caused by sampling error rather than coding bias.

- The risk-adjustment models estimated using reabstracted data have significantly greater discrimination than those estimated using original OSHPD data. However, they do not necessarily explain more of the variation in observed mortality across hospitals.

These findings suggest that as much as 29% of the difference in risk-adjusted mortality based on Model B, and as much as 12% of the difference based on Model A, may be attributable to variation in the coding of risk factors. In other words, Model B is somewhat compromised by coding bias but Model A is virtually immune. Even with Model B, however, at least 71% of the spread in risk-adjusted mortality is not explained by coding variation. These results are generally consistent with those obtained when the US Health Care Financing Administration's risk-adjustment approach was applied to data from an earlier, unblinded reabstraction study of medical-surgical DRGs in California.[6,7]

These results will not lead to any specific changes to the California Hospital Outcomes Project, but will be disseminated to hospitals as part of an ongoing effort to promote more complete and uniform coding of secondary diagnoses.

QUESTION 4: How often do the clinical characteristics used as risk factors in Model B actually represent conditions that developed after admission?

- Upon careful review of the timing of each diagnosis, risk factors fall into three groups:

    1. Conditions that are documented in ER or admission notes in less than 50% of cases and are first diagnosed at least one day after presentation in more than 50% of cases. Examples include hypotension, other cerebrovascular disease, pulmonary edema, other valve disease, and shock.

    2. Conditions that are documented in ER or admission notes in 50-80% of cases. Examples include congestive heart failure, chronic liver disease, complete atrioventricular block, epilepsy, secondary malignant neoplasm, nutritional deficiency, and skin ulcer.

---

[7] Green J, Wintfeld N. How accurate are hospital discharge data for evaluating effectiveness of care? *Med Care* 1993; 31:719-731.

3. Conditions that are documented in ER or admission notes in at least 80% of cases. Examples include infarct site, chronic renal disease, diabetes, hypertension, late effects of CVA, prior CABG, primary malignant neoplasm, and history of pacemaker. Many of these preexisting conditions are first noted on the day after presentation.

- The risk factors common to both Models A and B fall into the second and third categories, whereas most of the extra risk factors unique to Model B are in the first category. The major exceptions to this principle are: (1) other valve disease, which is often diagnosed during an inpatient echocardiogram or ventriculogram, and (2) complete atrioventricular block and epilepsy, which are actually present at admission in most cases. These findings generally support the manner in which risk factors were assigned to Models A and B.

- Adjusting only for pre-existing conditions compromises the discriminatory power of Model B more than that of Model A, although Model B remains stronger than Model A. It also substantially weakens both models' ability to explain the variation in observed mortality across hospitals.

- The regression coefficients for Model B risk factors are significantly biased by including conditions diagnosed after admission; epilepsy, hypotension, pulmonary edema, and shock become much less powerful predictors when this bias is removed.

- Disregarding conditions that were actually diagnosed after admission increases the difference in risk-adjusted mortality between low and high-mortality hospitals by 25% in a reestimated version of Model A, and by 20% in a reestimated version of Model B.

Very little comparative information on the timing of comorbid diagnoses has been published. After a "present on admission" indicator was implemented at the Mayo Clinic, Naessens et al reported that 77% of secondary diagnoses of cerebrovascular disease, 70% of secondary diagnoses of pneumonia, 64% of secondary diagnoses of acute renal failure, and 79% of secondary diagnoses of skin ulcer were preexisting.[8] Among AMI patients in New York, 63% of secondary diagnoses of urinary tract infection, 76% of secondary diagnoses of gastrointestinal hemorrhage, 71% of secondary diagnoses of cardiogenic shock, 66% of secondary diagnoses of acute renal failure, and 66% to 94% of secondary diagnoses of cerebrovascular disease were reported as "onset prior to admission."[9]

---

[8] Naessens JM, Brennan MD, Boberg CJ, Amadio PC, Karver PJ, Podratz RO. Acquired conditions: An improvement to hospital discharge abstracts. *Qual Assur Health Care* 1991; 3(4):257-263.

[9] Reepmeyer T. Complications versus comorbidities. Presented at the Faulkner and Gray Medical Outcomes Conference, Boston MA, July 1995.

This validation study shows that misclassifying conditions diagnosed after admission as risk factors leads to significant bias in the Model B regression coefficients. As a result, Model B over-adjusts for conditions that may represent complications of care and underestimates the true difference in risk-adjusted mortality between low-mortality and high-mortality outlier hospitals. This finding supports OSHPD's decision to report the results of the two models separately, and confirms the importance of adding a data element indicating whether each diagnosis was "present at admission" to the hospital discharge data system. It appears that reviewing prehospital, emergency room, and admission notes may be a more reliable way to ascertain these diagnoses than reviewing all notes written on the day of admission. However, there is clear potential for confusion when preexisting diagnoses are first detected during an inpatient diagnostic test.

Unfortunately, this new data element was just introduced in January 1996 and will not be available for use in California outcomes reports until late 1997. In the absence of a method to identify conditions that were actually present at admission, the misclassification bias described above will be resolved by dropping problematic variables (e.g., epilepsy, pulmonary edema, hypotension, and possibly shock) from all risk-adjustment models. OSHPD's third analysis of AMI mortality, now underway, will incorporate these changes.

QUESTION 5: How do the risk-adjustment models change when additional clinical variables are used as risk factors?

- Using both bivariate and multivariate statistical methods to test over 50 clinical risk factors abstracted from medical records, nine predictors that significantly improved the 1993 risk-adjustment models were identified. These predictors were divided into five core variables (i.e., systolic blood pressure, heart rate, and shock at presentation; cardiopulmonary arrest within 24 hours before presentation; and a "do not resuscitate" order on or before the date of admission) and four secondary variables (i.e., the ratio of the first CK to the hospital's upper limit of normal, pulmonary rales or a loud systolic murmur on the first physical examination; any history of stroke). The secondary variables either had marginal statistical ($0.03<p<0.10$) or clinical significance, or became insignificant when reabstracted ICD-9-CM codes were used instead of original OSHPD data.

- Adding clinical risk factors improves the discrimination of all risk-adjustment models, although the magnitude of this improvement is smaller for Model B than for Model A. The core clinical variables contribute much more than the secondary clinical variables, but the latter set of risk factors still improves the discrimination of most models. Adding clinical risk factors also increases both models' ability to explain the variation in observed mortality across hospitals.

- Although the magnitude of improvement from adding clinical variables is smaller when reabstracted ICD-9-CM codes are used in the "base" model instead of original codes, this finding reflects overadjustment for conditions that were actually diagnosed after admission. Limiting the analysis to risk factors that were clearly

present at admission, based on ER or admission notes, increases the magnitude of improvement from adding clinical variables.

The validation study clearly demonstrates that a better risk-adjustment model for AMI mortality could be developed if additional clinical information was available. The vital signs and presence or absence of shock at presentation, a recent history of cardiopulmonary arrest, and "do not resuscitate" status are the most important incremental predictors. Note that many potentially useful predictors could not be assessed, largely because the necessary data were not available with sufficient frequency (e.g., ejection fraction), were too costly to obtain (e.g., comparison ECGs before or after the index hospitalization) or were potentially related to the quality of hospital care (e.g., pre-infarct medications). Therefore, these variables would be inappropriate candidates for an enhanced statewide data collection program.

Several risk factors in Models A and B, such as insurance status and "other" infarct site, become much less significant when additional clinical variables are included. This finding reflects the greater explanatory power and precision of clinical variables, but does not necessarily indicate bias in how Models A and B estimate predicted probabilities. Instead, it demonstrates that the adjusted odds ratios reported in Chapter Ten of the 1993 report (and in Chapter Nine of this volume) should be interpreted cautiously when potential confounders are unavoidably omitted.

These results will not lead to any specific changes to the California Hospital Outcomes Project, but will be considered by the California Health Information Committee and the California Health Policy and Data Advisory Commission as part of their ongoing review of potential changes to OSHPD's data collection programs (pursuant to Senate Bill 1109).

QUESTION 6: Do hospitals with significantly higher or lower than expected mortality, as categorized in Volume One, appear closer to average after adjusting for additional clinical variables? How do the risk-adjusted mortality rates and p values for individual hospitals change when additional clinical variables are used as risk factors?

- In general, neither core nor secondary clinical variables systematically change expected mortality rates for hospitals with low, intermediate, or high risk-adjusted mortality.

- The addition of both core and secondary clinical risk factors to a reestimated version of Model A, based on the ICD-9-CM codes reported to OSHPD, reduces the difference in risk-adjusted mortality between low-mortality and high-mortality hospitals by 10%. The addition of the these risk factors to a similarly reestimated version of Model B reduces this difference by 20%.

- The addition of both core and secondary clinical risk factors to a reestimated version of Model A based on reabstracted ICD-9-CM data has a minimal effect on the difference in risk-adjusted mortality between low-mortality and high-mortality hospitals. The addition of these risk factors to a similarly reestimated version of

Model B reduces this difference by 21% if conditions diagnosed after admission are used in coding risk factors, and by 14% if they are not.

These findings demonstrate that unmeasured clinical risk factors account for little of the observed difference in risk-adjusted mortality across hospitals. Through sequential analysis of 925 cases that were included in all models, the relative contribution of various factors in "explaining" the observed difference in risk-adjusted mortality between low-mortality and high-mortality hospitals (using Model A) can be described as follows:

a. Random error (21%)
b. Bias due to differential coding of risk factor diagnoses (18%)
c. Bias due to differential timing of diagnoses (-16%)[10]
d. Bias due to unmeasured risk factors, or confounders (2%)

The net effect of these errors is that OSHPD's 1993 Model A overestimated the true difference in risk-adjusted mortality between these sets of hospitals by 24%.

Using Model B, the relative contributions are as follows:

a. Random error (22%)
b. Bias due to differential coding of risk factor diagnoses (19%)
c. Bias due to differential timing of diagnoses (-11%)[10]
d. Bias due to unmeasured risk factors, or confounders (10%)

The net effect of these errors is that OSHPD's 1993 Model B overestimated the true difference in risk-adjusted mortality between these sets of hospitals by 39%.

These analyses confirm that Model B provides a less valid portrayal of hospital performance than Model A, although a model that includes some but not all of the extra risk factors in Model B might be superior to either of the published models. If coding variation across hospitals can be eliminated, collecting and adjusting for additional clinical variables may have minimal impact at the hospital level. However, if coding variation remains unchanged, adjusting for additional clinical variables may be more important. Once again, this information will be considered by the California Health Information Committee and the California Health Policy and Data Advisory Commission as part of their ongoing review of potential changes to OSHPD's data collection programs. These committees are especially concerned about the potential impact of better risk models on the assessment of individual hospitals, as shown in Figures 14.3 and 14.4.

QUESTION 7: Do hospitals with low risk-adjusted mortality demonstrate better processes of care than hospitals with high risk-adjusted mortality?

- High volume hospitals administer aspirin to a higher percentage of AMI patients than medium-volume hospitals, but aspirin use does not differ across hospital mortality

[10] The negative sign indicates that the effect of this bias is to oppose the other biases listed; that is, it leads to underestimation of the true difference in risk-adjusted mortality.

categories. However, low-mortality hospitals start aspirin within 6 hours of presentation more often than intermediate or high-mortality hospitals.

- Low-mortality hospitals administer heparin to a higher percentage of AMI patients than intermediate or high-mortality hospitals, but there is no difference in heparin use between medium and high volume hospitals.

- Thrombolytic use is associated with neither hospital volume nor hospital mortality. This result is unaffected by whether a narrower or broader list of contraindications is used. Low-mortality and high-mortality hospitals also do not differ in the use of aspirin and heparin as early adjunctive therapy with thrombolytics.

- AMI patients admitted to low-volume hospitals are less likely to undergo PTCA, but are just as likely to undergo CABG, compared with those admitted to high-volume hospitals. Patients admitted to high-mortality hospitals are somewhat less likely to undergo CABG, but are almost as likely to undergo PTCA, compared with those admitted to low-mortality hospitals. Revascularization (CABG or PTCA) within 24 hours of presentation is about twice as frequent in low-mortality as in high-mortality hospitals.

- Coronary angiography and pulmonary artery (Swan-Ganz) catheterization are also performed more frequently at low-mortality than at high-mortality hospitals.

- There are no systematic differences in the measurable efficiency of emergency services between low-mortality and high-mortality hospitals.

The usage rates for these therapies in the validation sample are generally consistent with those recently reported for Medicare patients by the Cooperative Cardiovascular Project (CCP).[11] For example, aspirin was used in 83% of "ideal candidates" in the 1992-93 CCP sample from Alabama, Connecticut, Iowa, and Wisconsin, versus 73% of those eligible in OSHPD's validation sample. Thrombolytics were used in 70% of "ideal candidates" in the CCP sample, versus 51% of those eligible in OSHPD's validation sample. Intravenous or subcutaneous heparin was used in 69% of "ideal candidates" in the CCP study, versus 63% of those eligible in the validation sample.

These analyses indicate that there are definite differences in the process of care between hospitals with low risk-adjusted mortality and hospitals with high risk-adjusted mortality, stratified by volume. The most clinically significant differences relate to the use of invasive diagnostic and therapeutic methods, including catheterization, PTCA, and CABG. However, low-mortality hospitals demonstrate good outcomes even among their patients who do not receive one of these invasive procedures. This finding is consistent with the hypothesis that procedure use is one marker of "aggressiveness" that may be associated with better AMI outcomes; other markers of this trait probably exist but could not be identified through retrospective chart review.

---

[11] Ellerbeck EF, Jencks SF, Radford MJ, Kresowik TF, Craig AS, Gold JA, Krumholz HM, Vogel RA. Quality of care for Medicare patients with acute myocardial infarction. *JAMA* 1995; 268:2530-2536.

In summary, the results of this validation study demonstrate that:

1. False-positive coding of AMI is much less frequent in California in the 1990's than earlier studies would suggest, and does not compromise the validity of the California Hospital Outcomes Project.

2. Although many risk factors are undercoded, there is no systematic difference in coding practice between hospitals with low risk-adjusted mortality and hospitals with high risk-adjusted mortality.

3. The difference in risk-adjusted mortality between low-mortality and high-mortality hospitals decreases modestly when uniformly coded data are substituted for the data originally reported to OSHPD.

4. Clinical risk factors available only through chart review, such as vital signs, shock, cardiopulmonary resuscitation, and "do not resuscitate" status, significantly improve the performance of risk-adjustment models based on ICD-9-CM data.

5. Adjusting for these clinical risk factors has a minor impact on hospitals' risk-adjusted mortality rates.

6. Model B suffers from more bias than Model A, largely because it includes risk factors that are frequently diagnosed after admission and occur more often at high-mortality hospitals than at low-mortality hospitals.

7. Process-of-care differences between low-mortality and high-mortality hospitals can be identified, but do not fully explain the observed differences in risk-adjusted outcomes (even after adjustment for clinical risk factors).