A Statistical Framework for Severity Adjustment of Hospital Mortality Rates

Neal Thomas, Nicholas T. Longford, John E. Rolph
The research described in this report was sponsored by the Health Care Financing Administration, U.S. Department of Health and Human Services, under Cooperative Agreement No. 99-C-98489/9-06.

RAND is a nonprofit institution that seeks to improve public policy through research and analysis. Publications of RAND do not necessarily reflect the opinions or policies of the sponsors of RAND research.
A Statistical Framework for Severity Adjustment of Hospital Mortality Rates

Neal Thomas, Nicholas T. Longford, John E. Rolph

Prepared for the Health Care Financing Administration, U.S. Department of Health and Human Services
PREFACE

The research leading to this Note was supported by the Health Care Financing Administration (HCFA) through cooperative agreement 99-C-98489/9-06 with the RAND/UCLA/Harvard Center for Health Care Financing Policy Research and by RAND’s own funds.
SUMMARY

We present a statistical framework and associated data analyses that should inform the interpretation of hospital death rates for Medicare patients. The hospital-level 30-day survival rates for Medicare pneumonia patients in fiscal year 1986, for example, varied from 0 percent to 100 percent among the nation’s 6000 acute care hospitals. This tremendous variation begs for an explanation. Candidate causes of this variation include small sample fluctuations, mix in severity of patient condition (“patient severity”), and hospital quality of care. In our statistical modeling and analyses, we attempt to characterize how much of this variation can be attributed to each of these causes.

MORTALITY DATA FOR ASSESSING HOSPITAL QUALITY

The Medicare Mortality Predictor System (MMPS) collected a nationally representative sample of 5888 hospital discharges for the four medical conditions stroke, pneumonia, myocardial infarction, and congestive heart failure. The MMPS is a computerized medical record abstraction system that allows each hospital to compare its raw mortality rate for each of the four diseases to an adjusted rate that estimates the national mortality rate for patients with similar clinical conditions to those admitted by the hospital.

Since the sampling design for the MMPS had too few patients per hospital to accurately estimate the variation in the hospital-level severity adjusted mortality rates, we used data from the DRG-based Prospective Payment System (PPS) RAND quality of care study, which has more patients per hospital than the MMPS. The PPS study includes 60 to 80 disease-specific variables from the medical record that describe the patient’s acute and chronic, morbid and comorbid conditions. We use logistic regression to account for differences in patient mix when adjusting hospital-specific mortality rates. We also include a component of variance to account for the hospital-to-hospital variation in underlying death rates that cannot be accounted for by differences in patient severity.

REPORTING OF HOSPITAL-SPECIFIC MORTALITY DATA

Since the observed hospital mortality rate in a given year reflects the underlying rate plus or minus random variation from the individual patients, it is important (1) to produce accurate standard errors of individual hospital rates and (2) to make an overall assessment of how much the underlying hospital-specific mortality rates vary across all hospitals. The Health Care Financing Administration (HCFA) reports severity-adjusted individual hospital mortality rates and associated standard errors based on administrative data, and the MMPS
provides for each hospital collecting its own data and producing adjusted death rates and standard errors by disease category. We compare alternative estimators of hospital-specific mortality rates and their standard errors to the currently used ones with a simulation study. The rationale for reporting differences (with or without severity adjustment) in hospital-specific mortality rates is the assumption that they are more than just noise from sampling fluctuations. One contribution of this study is putting observed mortality rates into a statistical framework that allows us to estimate how much hospitals vary in their underlying mortality rates both with and without severity adjustment. We also present measures of how accurately we can estimate this underlying variation. Both the point estimates of the amount of variation and their accuracy bear on how one should interpret hospital-specific mortality rates.

Do Hospitals Differ in Their Underlying Mortality Rates?

It is abundantly clear from national data that hospitals differ in their underlying death rates after accounting for sampling variation but not for patient severity. Our variance component models confirm this finding across the same four medical conditions of stroke, pneumonia, myocardial infarction, and congestive heart failure. Our logit-normal variance component models using the PPS data confirm earlier findings that hospital mortality rates do vary with patient severity. However the severity-adjusted extrabinomial variation in hospital death rates is poorly estimated, and the data suggest differing amounts of variation in mortality rates for the four medical conditions. Indeed, only the congestive heart failure variance components can be distinguished from zero by statistical significance tests. How to combine the variance components of the national data with no severity adjustment with the PPS data with severity adjustment in a consistent way is unclear.

Implications for Individual Hospital Reporting

HCFA currently releases reported Medicare death rates with standard errors by hospital for 17 diagnostic categories from national data. It also supplies hospitals, through the MMPS, with a way to calculate their own severity-adjusted death rates and corresponding standard errors. Our Bayesian derivation of the MMPS estimator and its standard error is based on a diffuse prior distribution for the hospital effects. Contrasting this diffuse prior distribution to our estimates of the variance component of the severity-adjusted hospital effects gives an interpretable comparison of the MMPS estimator and the Bayes shrinkage estimators of hospital death rates. The MMPS estimator through the flat prior distribution ignores the information in the PPS data showing that individual hospital effects cluster. This suggests that estimates of the death rate for a single hospital might be
improved by using Bayes estimators. The idea here is that an individual hospital's observed death rate may be extreme through a combination of having an extreme underlying rate and having large sampling variation. The shrinking of the Bayes estimator adjusts for the sampling variation component.

Our simulation comparisons of the MMPS procedure and the Bayes-based procedures for hypothetical hospital populations reveal some interesting conclusions. As expected, the Bayes shrinkage estimator performs relatively well for hospitals with death rates near the national norm and relatively less well for more extreme values.

The Bayes estimator produces biased estimates of hospital differences and has a lower variance than the MMPS estimator. By shrinking large estimates of hospital differences strongly toward zero, the increased stability of the Bayes estimator gives better estimates for hospitals that have extremely variable death rates that are not demonstrably extreme. An important use of these estimators is in identifying exceptional hospitals. The Bayes procedure is more conservative in assessing the evidence that a hospital is extreme than the MMPS estimator, although the two procedures perform similarly (and rather poorly) in identifying extreme hospitals.

NEXT STEPS

By incorporating information about the population of severity-adjusted hospital death rates into a statistical framework for estimating individual hospital severity-adjusted death rates, we have moved forward in improving the available estimation methodology. However, the PPS design provided such a thin sample per hospital that we were not able to definitively address how much hospitals vary in their underlying death rates after adjustment for patient condition. To address this question a database with patient detail like the PPS or the MMPS, but with many more patients per hospital, would need to be assembled and analyzed. One possible source of such data may be by compromising on the detail of patient severity information and using the national administrative data.
ACKNOWLEDGMENTS

We acknowledge the role of David Draper in writing the proposal for funding this work and his assistance with some data processing tasks. Ellen Harrison and Daniel Relles made substantial contributions to the data management necessary to carry out this research. Charles Lewis and Donald Rubin provided helpful advice about the numerical evaluation of Bayes shrinkage estimates. Emmett Keeler’s excellent review of an earlier draft led to a much improved presentation.
CONTENTS

PREFACE ................................................................. iii
SUMMARY ............................................................... v
ACKNOWLEDGMENTS ................................................. ix
FIGURES ............................................................... xiii
TABLES ............................................................... xv
SYMBOLS ............................................................. xvii

Section
1. INTRODUCTION .................................................. 1
Mortality Data for Assessing Hospital Quality ....................... 1
Research Issues Addressed by the Proposed Methodology ........ 1
Roadmap ............................................................. 2

2. DATA SOURCES .................................................. 4
MMPS and PPS Data .................................................. 4
Patient Severity Measures ........................................... 4
MMPS National Sample ............................................. 5
PPS Evaluation Sample ............................................ 5
Medicare National Mortality Data .................................... 6
Hospital-Specific MMPS Data ....................................... 7

3. STATISTICAL METHODS .......................................... 8
Statistical Formulation ............................................... 8
Bayes Shrinkage Estimators ......................................... 12
A Bayesian Derivation of the Estimator of Hospital Mortality Currently Used
by the MMPS ......................................................... 13

4. RESULTS ........................................................... 16
Observed Death Rates Without Severity Adjustment .............. 16
Estimates of the Variance Components Without Severity Adjustment Based on
the Medicare National Mortality Data ................................ 16
Estimates of the Variance Components with Severity Adjustment Based on the
PPS Data ............................................................ 16
Combining Information About Variance Component Estimates from Different
PPS Data Sources .................................................... 19
Constructing a Prior for \( \delta \) ............................................ 19

5. SIMULATION STUDIES COMPARING THE SHRINKAGE ESTIMATOR
AND THE ESTIMATOR CURRENTLY USED BY THE MMPS .......... 22
Overview of Simulation Results ...................................... 22
Design of the Simulation Study with Repeated Sampling from a Small Number
of Hypothetical Hospitals ........................................... 24
Design of the Simulation Study with Repeated Sampling from Fixed Hospital
Settings ............................................................. 25
Simulation Estimates of the Statistical Properties of the Estimators .... 26

6. DISCUSSION ....................................................... 36
Do Hospitals Differ in Their Underlying Mortality Rates? ........ 36
Appendix
A. STANDARD ERRORS BASED ON A DIFFUSE PRIOR DISTRIBUTION ........ 39
B. STANDARD ERRORS BASED ON REPEATED SAMPLING
   CONSIDERATIONS .......................................................... 44
C. SIMULATION STUDY OF STANDARD ERRORS FOR $\bar{Y} - \hat{P}_{N_{mle}}$ ............ 49
D. SAMPLE DESIGN FOR A FUTURE MMPS ............................................. 57
E. COMPARING THE INFORMATION FROM THE NATIONAL
   DISTRIBUTION WITH THE HOSPITAL-SPECIFIC INFORMATION ........... 58
F. COMPUTING MAXIMUM LIKELIHOOD ESTIMATES FOR THE LOGIT-
   NORMAL MODEL ............................................................... 61
G. EVALUATING THE MARGINAL POSTERIOR DISTRIBUTION OF $\delta$ AND
   $\bar{P}_{H} - \bar{P}_{N}$ ....................................................................... 63
H. SIMULATING NATIONAL MAXIMUM LIKELIHOOD ESTIMATES ...................... 66
I. CONSTRUCTING HYPOTHETICAL HOSPITALS WITH PATIENTS
   HAVING DIFFERENT MIXES OF SEVERITY ................................... 67
J. DERIVATION OF THE EQUIVALENCE OF $\bar{P}_{H} - \bar{P}_{N}$ AND $\bar{Y} - \bar{P}_{N_{mle}}$ .......... 69
BIBLIOGRAPHY ............................................................................. 71
FIGURES

4.1. Histogram of the Death Rate for Pneumonia Observed at Each Hospital (a), Histogram of Simulated Hospital Death Rates Assuming No Differences Between Hospitals (b), and Difference Between These Two Histograms (c) .............................................. 17

4.2. Minus Twice the Log Likelihood Function for $\sigma_\delta$ Evaluated at the Maximum Likelihood Estimates of the Regression Parameters ............................................. 20

5.1. Coverage Probabilities (Using $\overline{\bar{Y}} - \overline{\bar{P}}_{N,mle}$ and $se_c$) Compared to the National Death Rate for Patients Like Those Treated at the Hospital ($\overline{\bar{P}}_N$) .............................................. 29

5.2. Histogram of the 50 Most Extreme Values of $\bar{P}_H - \bar{P}_N$ from the Simulation Sample of 2000 (a) and Histogram of the 50 Values of $\bar{P}_H - \bar{P}_N$ Selected Using the Current MMPS Z-statistic (b) .............................................. 34

A.1. The Prior Distribution for $\delta$ Implied by the Imputation of an Observation into the Likelihood Function ................................................................................. 42

C.1. Plots of $\text{var}\left(\overline{\bar{P}}_{N,mle}\right)/\text{var}\left(\overline{\bar{Y}} - \overline{\bar{P}}_{N,mle}\right)$ by Hospital Sample .................................................. 52

C.2. The Difference Between the Average (Across Simulation Replications) Variance Computed for $se_u$ and the Population-Based Variance vs the Population Variance .................................................................................. 54

C.3. Coverage Probabilities Estimated from the Simulation Using $se_u$ vs the Hospital Death Rate ........................................................................................................ 54

C.4. Coverage Probabilities Estimated for Each Simulation Setting Computed Using $se_{nu}$ vs the National Death Rate for Patients Like Those Treated at the Hospital .................................................................................. 55

C.5. Coverage Probabilities Estimated for Each Simulation Setting Computed Using $se_{b1}$ vs the National Death Rate for Patients Like Those Treated at the Hospital .................................................................................. 56
### TABLES

2.1. Severity Characteristics Used in the MMPS ........................................ 5
2.2. Quantiles of the Hospital Medicare Sample Size Distributions for 1986 ........ 7
4.1. Variance Component Estimates Without Severity Conditioning Based on the National Database for 1986 ................................................................. 18
4.2. Variance Component Estimates with Severity Adjustment Based on the PPS Data .......................................................... 18
5.1. Coverage Probabilities When $\delta$ Are Randomly Generated .................. 27
5.2. Coverage Probabilities for Nominal 95 Percent Bayes-Based Intervals for Each $\delta$ Value Averaged Across the Simulation Settings ......................... 28
5.3. Ratio of MSE of $\bar{y} - \bar{P}_{N_{mle}}$ and MSE of $\bar{P}_{H_0} - \bar{P}_{N_{b}}$ .......................................................... 31
5.4. Relative Bias, $\{E(\bar{P}_{H_0} - \bar{P}_{N_{b}})-(\bar{P}_H - \bar{P}_N)\}/\sqrt{\bar{P}_H - \bar{P}_N}$, of the Bayes Procedure Averaged Across Medical Conditions and Severity .......................... 31
5.5. Estimates of Power for Testing the Null Hypothesis $\bar{P}_H - \bar{P}_N = 0$ at .05 Level .................................................. 32
5.6. The Mean of $\bar{P}_H - \bar{P}_N$ for Hospitals Identified as Extremely Low or High Quality by the Different Ranking Criteria ........................................... 33
C.1. Summary Statistics from All Simulation Settings .................................... 50
C.2. Summary Statistics from Simulation Settings with $n = 25$ ..................... 50
C.3. Summary Statistics from Simulation Settings with $n = 25$ and Lowest Patient Severity .................................................. 51
C.4. Estimates of Power for the Null Hypothesis Using Different Standard Error Estimates ........................................ 52
E.1. The Asymptotic Variance, $\sigma^2_\delta$, of the Maximum Likelihood Estimate of $\delta$, with Each Entry Multiplied by 200 for a Similar Hospital of Size $n$, Table Entry = $N\sigma^2_\delta$ .................................................. 59
E.2. The Sample Size at a Hospital Required to Match the Prior Information at the Hospital, with the Range Given for Each Entry Based on the Values $\hat{\delta}_{mle} = -0.30, 0.30$ .................................................. 60
SYMBOLS

\( \bar{P}_N \) The average probability of death in a hospital computed using the national death rate for patients with the same severity.

\( P_{N,i} \) The probability of death for the \( i \)th patient in a hospital based on the death rate for similar patients in the entire nation.

\( \bar{P}_H \) The average probability of death in a hospital computed using the hospital-specific death rate.

\( P_{H,i} \) The probability of death for the \( i \)th patient in a hospital based on the hospital-specific death rate.

\( \alpha \) The intercept in the logistic regression death rate models.

\( \beta \) The slope in the logistic regression death rate models.

\( \delta \) The hospital-specific increase (decrease) in the death rates compared to the national death rates.

\( \sigma_\delta^2 \) The variance of \( \delta \) measuring the variation in quality across hospitals.

\( \hat{\alpha}_{mle}, \hat{\beta}_{mle} \) Estimates of \( \alpha \) and \( \beta \) based on the MMPS data.

\( \hat{\Sigma}_{mle} \) The asymptotic covariance matrix of \( (\hat{\alpha}_{mle}, \hat{\beta}_{mle}) \).

\( \hat{\alpha}, \hat{\beta} \) Updated maximum likelihood estimates of \( \alpha \) and \( \beta \) using the data available from individual hospitals.

\( \hat{\delta} \) Maximum likelihood estimate of \( \delta \).

\( \hat{\Sigma} \) The asymptotic covariance matrix of \( (\hat{\alpha}, \hat{\beta}, \hat{\delta}) \).

\( \hat{\Sigma}_\delta \) The vector of covariances of \( \delta \) with \( (\alpha, \beta) \) in the Bayesian calculations.

\( \hat{\alpha}_{b1}, \hat{\beta}_{b1}, \hat{\delta}_{b1} \) Bayes estimates of \( (\alpha, \beta, \delta) \) based on the one-observation prior.

\( \hat{\Sigma}_{b1} \) The asymptotic covariance matrix of \( (\hat{\alpha}_{b1}, \hat{\beta}_{b1}, \hat{\delta}_{b1}) \).

\( \hat{\alpha}_b, \hat{\beta}_b, \hat{\delta}_b \) Bayes estimates of \( (\alpha, \beta, \delta) \) based on the informative prior.

\( \hat{\Sigma}_b \) The asymptotic covariance matrix of \( (\hat{\alpha}_b, \hat{\beta}_b, \hat{\delta}_b) \).

\( \hat{\alpha}_{sim}, \hat{\beta}_{sim}, \hat{\Sigma}_{sim} \) Simulated values of \( \hat{\alpha}_{mle}, \hat{\beta}_{mle}, \hat{\Sigma}_{mle} \).

\( \bar{P}_{N,mle}, \bar{P}_{N,mle,i} \) Estimates of \( \bar{P}_N \) and \( P_{N,i} \) based on the maximum likelihood estimates.

\( \bar{P}_{H,mle}, \bar{P}_{H,mle,i} \) Estimates of \( \bar{P}_H \) and \( P_{H,i} \) based on the maximum likelihood estimates.

\( \bar{P}_N, \bar{P}_{N,i} \) Estimates of \( \bar{P}_N \) and \( P_{N,i} \) based on the updated maximum likelihood estimates.

\( \bar{P}_H, \bar{P}_{H,i} \) Estimates of \( \bar{P}_H \) and \( P_{H,i} \) based on the updated maximum likelihood estimates.
Estimates of $\hat{P}_{N}$ and $P_{N,i}$ based on the one-observation Bayes prior.

Estimates of $\hat{P}_{H}$ and $P_{H,i}$ based on the one-observation Bayes prior.

Estimates of $\hat{P}_{N}$ and $P_{N,i}$ based on the informative Bayes prior.

Estimates of $\hat{P}_{H}$ and $P_{H,i}$ based on the informative Bayes prior.

$\hat{t}_i$ The estimated logit for the $i$th patient in the hospital sample using the national death rate model and the MLEs obtained from the MMPS sample.

$se_c$ The "conditional" standard error estimate of $\bar{Y} - \bar{P}_{n_{mm}}$.

$se_u$ The "unconditional" standard error estimate for $\bar{Y} - \bar{P}_{n_{mm}}$.

$se_{mle}$ The standard error estimate for $\bar{Y} - \bar{P}_{n_{mle}}$ based on a diffuse prior.

$se_b$ Standard error estimate for $\bar{P}_{H} - \bar{P}_{N}$.

$se_{b1}$ Standard error estimate for $\bar{P}_{H1} - \bar{P}_{N1}$.

$X_i$ The covariates measuring severity for the $i$th patient in a sample of patients from a hospital.

$Y_i$ A 0/1 variable that equals one if the $i$th patient in a sample of patients from a hospital died within 30 days.

$n$ Number of patients in the sample from a hospital whose death rate is being adjusted.

$X_i^m$ The severity characteristics of the $i$th patient in the MMPS sample.

$Y_i^m$ A 0/1 indicator of 30-day survival for the $i$th patient in the MMPS sample.

$n_m$ Number of patients in the MMPS sample.

$X_j^P$ The severity characteristics of the $j$th patient in the $i$th hospital in the PPS sample.

$Y_j^P$ A 0/1 indicator of 30-day survival for the $j$th patient in the $i$th hospital in the PPS sample.

$n_1, \ldots, n_P$ The number of patients at each hospital in the PPS data set.

$n_1, \ldots, n_T$ The number of patients in each hospital during fiscal year 1986.
1. INTRODUCTION

In this Note, we present a statistical framework and associated data analyses that should inform the interpretation of hospital death rates for Medicare patients. The hospital-level 30-day survival rates for Medicare pneumonia patients in fiscal year 1986, for example, varied from 0 percent to 100 percent among the nation's 6000 acute care hospitals. This tremendous variation begs for an explanation. Possible causes of this variation include small sample fluctuations, hospital differences in condition severity, and quality of care. In the statistical modeling and analyses reported here, we attempt to characterize how much of this variation can be attributed to each of these causes.

MORTALITY DATA FOR ASSESSING HOSPITAL QUALITY

Release of hospital-level mortality data by the Health Care Finance Administration (HCFA), state organizations, and private groups in recent years has focused attention on how hospital-level mortality rate data can be used to evaluate the effectiveness of hospital care. There is some controversy about the value of hospital mortality data for measuring the quality of care (Kahn et al., 1988a; Dubois et al., 1987). The principal objection is that mortality rates obtained from administrative data sources are difficult to interpret because there has been little adjustment for the clinical condition of patients on admission (Green et al., 1990).

We use data from the Medicare Mortality Predictor System (MMPS) (Daley et al., 1988; Jencks et al., 1988), which collected a specially selected sample of patients and hospitals, to enable us to adjust for the mix of severity of patient conditions ("patient severity"). These data contain the best currently available severity information for adjusting mortality rates. However, this Note focuses on the statistical methodology used to report hospital-specific mortality rates. The use of the MMPS data is intended to illustrate how these statistical methods, with minor modifications, could be used with any reporting system that releases hospital-specific mortality rates adjusted for differences in patient mix.

RESEARCH ISSUES ADDRESSED BY THE PROPOSED METHODOLOGY

A hierarchical Bayesian statistical model is used to represent the different sources of variation in observed hospital-specific mortality rates. This fully parametric approach provides a coherent framework for evaluating the relative importance of different sources of variation in the observed death rates. The fully parametric Bayesian framework contains as a special case the estimator now commonly used in mortality reporting systems (for example,
Health Care Finance Administration, 1989). It also provides standard inferential procedures for computing confidence or high probability coverage intervals, in contrast to existing methods, which have not been derived within a rigorous analytic framework. Several of the benefits of this approach are listed below.

1. The fully model-based formulation provides a principled, widely accepted standard error for the estimator currently in common use. Simulation study shows that the model-based standard error performed modestly better than several alternative standard errors in current use.

2. The variation in hospital-specific mortality rates not attributable to small sample fluctuations or measurable differences in patient severity is explicitly quantified. Previous studies, such as Park et al., 1990), have focused on establishing the existence of such variation.

3. We can evaluate the value, in terms of knowledge gained, of reporting the hospital-specific death rates based on different sample sizes. For example, how much more can we expect to learn about a hospital’s underlying death rate from a sample of 50 of its patients beyond what the national distribution of hospital mortality rates (adjusted for patient severity differences) tells us? Results of this type could be used for establishing minimum sample sizes for reporting purposes. In some cases, these calculations may discourage the collection of expensive severity information if the sample sizes at a hospital are too small to meaningfully improve our knowledge about the hospital.

4. The Bayesian methods provide more accurate, stable estimates of univariate, medical-condition-specific mortality rates.

5. The confidence or high probability coverage intervals produced by the Bayesian methods are corrected for the severe “multiple comparisons” problem that arises when “exceptional” hospitals are identified from among thousands of hospitals with the techniques in current use.

6. The hierarchical Bayesian models can be extended in a natural fashion to improve the accuracy of reporting by appropriately pooling across several medical conditions or time periods.

ROADMAP

In Section 2 we describe the three relevant data sources for the study: the MMPS, the PPS, and the Medicare national mortality data. Section 3 covers statistical methods. It uses a Bayesian approach to produce new “shrinkage” estimators of individual hospital mortality rates and to give a valid derivation of the current MMPS hospital mortality rate estimator.
Section 4 presents results on the variation in hospital-specific mortality rates with and without adjustment for patient severity. In Section 5, we report the results of a simulation study comparing how well the Bayesian shrinkage estimators and the MMPS estimator perform in estimating individual hospital mortality rates. The appendixes contain technical derivations and cover various side issues.
2. DATA SOURCES

MMPS AND PPS DATA

The MMPS (Daley et al., 1988; Jencks et al., 1988) collected a nationally representative sample of 5888 hospital discharges for the four medical conditions stroke, pneumonia, myocardial infarction, and congestive heart failure. The MMPS is a computerized medical record abstraction system that allows each hospital to compare its own raw mortality rate for each of the four diseases to an adjusted rate based on the estimated national mortality rate for patients with similar clinical conditions to those admitted by the hospital. The clinical variables used in the MMPS are described in Daley et al. (1988) and they are briefly summarized below.

Since the sampling design for the MMPS had too few patients per hospital to accurately estimate the variation in the hospital-level severity adjusted mortality rates, we used data from the DRG-based PPS RAND quality of care study (Kahn et al., 1990a). The PPS sampling plan’s advantage is more patients per hospital than the MMPS. The PPS design allows us to more accurately assess the hospital-level mortality rate variation that remains after adjustment for severity. Both the MMPS and PPS collect similar severity information.

PATIENT SEVERITY MEASURES

To interpret mortality rates as a measure of a hospital’s quality of care, we must know the patient’s condition at admission. Hospital-specific mortality rates can then be adjusted for differences in the mix of patient severity. The PPS designed a collection method for patient severity at admission that was used for the MMPS and PPS studies (Kahn et al., 1988b; Daley et al., 1988). It includes 60 to 80 disease-specific variables from the medical record that describe the patient’s acute and chronic, morbidity and comorbid conditions (Kahn et al., 1988b; Kosecoff et al., 1988; Roth et al., 1988; Rubenstein et al., 1988; Sherwood et al., 1988). We specify the form of the severity adjustment in our hospital mortality models from earlier models developed for the MMPS (Daley et al., 1988) and the PPS (Keeler et al., 1990).

The analyses in this Note make use of several different data sources. We briefly describe each data source in this section. Names and notation are introduced here to clarify references to the data sources.
MMPS NATIONAL SAMPLE

A large nationally representative sample of patients was collected for each of the four medical conditions in the MMPS. Mortality outcomes (30 days from hospital admission) for the \( n_m \) patients in the sample are represented by \( Y^m_i, \ldots, Y^m_{n_m} \), and their severity characteristics are denoted by \( X^m_i, \ldots, X^m_{n_m} \). To distinguish the patients in this sample from patients in a hospital whose death rate is being adjusted, their mortality outcomes and severity characteristics include the superscript \( m \). Our notation does not distinguish between the four medical conditions studied here (stroke, pneumonia, myocardial infarction, and congestive heart failure). Each medical condition is treated separately in the analyses. A brief description of the severity characteristics is given in Table 2.1. A detailed description of the use of the variables in the severity adjustment models is given in Daley et al. (1988).

The most common use of the MMPS data in this Note is through the maximum likelihood estimates of the parameters of the national survival rate function (Eq. 3.4), and their asymptotic covariance matrix. The severity characteristics of subsets of the MMPS data are also used to form the simulated hospitals presented in Section 5.

PPS EVALUATION SAMPLE

The patients sampled for the MMPS project were selected from many different hospitals, so that few hospitals included in the MMPS hospitals had more than two patients

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>APII</td>
<td>Apache II, commercial severity adjustment index (Knaus, Draper, and Wagner, 1985)</td>
</tr>
<tr>
<td>MALE</td>
<td>One if patient is male</td>
</tr>
<tr>
<td>AGE</td>
<td>Age (must be at least 65)</td>
</tr>
<tr>
<td>WALK</td>
<td>One if patient unable to walk</td>
</tr>
<tr>
<td>CANCER</td>
<td>Coexisting metastatic cancer</td>
</tr>
<tr>
<td>DNR</td>
<td>Do not resuscitate order in effect</td>
</tr>
<tr>
<td>SUN</td>
<td>Serum urea nitrogen</td>
</tr>
<tr>
<td>TEMP</td>
<td>Temperature</td>
</tr>
<tr>
<td>MAP</td>
<td>Mean arterial (blood) pressure</td>
</tr>
<tr>
<td>GCS</td>
<td>Glasgow Coma Score (stroke patients only)</td>
</tr>
<tr>
<td>SCAN</td>
<td>Mass effect on CT or MRI scan (stroke patients only)</td>
</tr>
<tr>
<td>PBC</td>
<td>One if positive blood culture (pneumonia patients only)</td>
</tr>
<tr>
<td>LOBE</td>
<td>One if more than one lobe involved on roentgenogram (pneumonia patients only)</td>
</tr>
<tr>
<td>SUBINF</td>
<td>One if subendocardial infarction (myocardial infarction only)</td>
</tr>
<tr>
<td>CHFRCR</td>
<td>One if congestive heart failure on chest roentgenogram (myocardial infarction only)</td>
</tr>
<tr>
<td>HRATE</td>
<td>Heart rate, beats/minute (myocardial infarction only)</td>
</tr>
<tr>
<td>HEMO</td>
<td>Hematocrit (congestive heart failure only)</td>
</tr>
</tbody>
</table>
represented. Although this is a good sampling design for the purpose of estimating severity adjustments, it is a very poor design for estimating how much of the nonsampling variability between hospitals can be attributed to differences in the severity of the patients they treat.

Because of their more appropriate design, the PPS data (Kahn et al., 1990a) were used here. The PPS project evaluated the change in the quality of care resulting from the introduction of the DRG-based PPS. The PPS evaluation studied the same medical conditions as the MMPS project, plus two additional medical conditions not studied in this Note. The PPS study and the MMPS project share the same data abstraction instruments, which were developed by the PPS study (Kahn et al., 1988b). (A few minor differences in the data collection instruments in the MMPS and PPS forced us to use variables with slightly different definitions in a few cases.)

Here is a very brief description of the sampling plan employed by the PPS project (see Draper et al. (1990) for details). Approximately nine patients were sampled for each medical condition from each of 297 hospitals. The number of patients at each hospital varied slightly because of the lack of eligible patients at some hospitals. Two features of the sampling design cause complications for the analyses in this Note. The first is that the patients from each hospital were selected from two time periods: half of the patients were admitted during 1981/82, before the introduction of the PPS, and half were admitted during 1985/86, after the introduction of the PPS. The second complication is that hospitals were selected through a sampling plan that stratified on hospital variables representing hospital size, urbanicity, and hospital poverty rate. Very small hospitals were also excluded from the sample. The implications of this sampling design are discussed in Section 3 where the PPS data are used to estimate the differences in hospital death rates after adjustment for severity differences among patients.

The PPS data are distinguished from the other datasets used in this Note by a superscript $p$ on the survival and severity variables, $Y_{ij}^p$ and $X_{ij}^p$, where $i$ denotes the hospital, and $j$ refers to the patients within each hospital. The number of patients sampled at each hospital in the PPS study is denoted by $n_{1p}, \ldots, n_{297p}$.

**MEDICARE NATIONAL MORTALITY DATA**

Hospital-specific mortality data (from approximately 5500 hospitals) for the entire population of Medicare patients in each of the four MMPS medical conditions for fiscal year 1986 are used to give accurate estimates of the heterogeneity in hospital-specific death rates for each condition. The observed national death rate at the $i^{th}$ hospital is denoted by $\bar{Y}_i^T$, and the number of patients at the $i^{th}$ hospital is denoted by $n_i^T$, $i = 1, \ldots, T$. 
Even if all hospitals provide the same quality of care and treat patients with identical severity, the actual annual death rates would vary because of the small number of qualifying patients at each hospital. Table 2.2 summarizes the distribution of the number of Medicare patients in each of the four medical conditions treated at each acute care hospital in the United States during fiscal year 1986. This database is analyzed in Jencks et al. (1988).

Table 2.2
Quantiles of the Hospital Medicare Sample Size Distributions for 1986

<table>
<thead>
<tr>
<th>Condition</th>
<th>10%</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
<th>90%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stroke</td>
<td>5</td>
<td>14</td>
<td>33</td>
<td>73</td>
<td>122</td>
</tr>
<tr>
<td>Pneumonia</td>
<td>13</td>
<td>28</td>
<td>56</td>
<td>99</td>
<td>153</td>
</tr>
<tr>
<td>Myocardial infarction</td>
<td>5</td>
<td>12</td>
<td>30</td>
<td>66</td>
<td>113</td>
</tr>
<tr>
<td>Congestive heart failure</td>
<td>10</td>
<td>23</td>
<td>54</td>
<td>111</td>
<td>188</td>
</tr>
</tbody>
</table>

HOSPITAL-SPECIFIC MMPS DATA

An important feature of the MMPS is that each hospital collects its own data voluntarily with no centralized reporting. As a consequence, the MMPS reports from hospitals cannot reference or use the MMPS results from other hospitals. Because there is currently no centralized reporting, we do not have hospital-specific MMPS data. The mortality data collected at a hospital applying the MMPS procedures are represented using the same \((X_i, Y_i), i = 1, \ldots, n\) notation without any superscripts.
3. STATISTICAL METHODS

STATISTICAL FORMULATION

Notation

This section presents a statistical framework for quantifying how the 30-day mortality rates vary across hospitals. The severity information on an individual patient is denoted by $X$, and $Y$ is a dichotomous variable taking the value 1 if the patient dies within 30 days of hospital admission and 0 otherwise. The proportion of patients who die among all patients in the nation with the same severity, $X$, is denoted by $P_N(Y = 1 | X)$. The $N$ subscript indicates that this rate is for all patients with the same severity in the entire nation. This quantity might be thought of either as the actual death rate in the nation for a given condition during a given year or as an unobservable probability of death averaged over the nation's hospitals and patient population.

The MMPS used a linear logistic regression model (Daley et al., 1988),

$$\text{logit}(P_N(Y = 1 | X)) = \alpha + \beta X,$$  \hspace{1cm} (3.1)

to represent the national death rate for patients with severity characteristics, $X$, for a given medical condition. They computed maximum likelihood estimates of the ($\alpha, \beta$) parameters using the large national sample collected by the MMPS. The maximum likelihood estimates are denoted by ($\hat{\alpha}_{mle}, \hat{\beta}_{mle}$), and the asymptotic variance-covariance matrix is denoted by $\hat{\Sigma}_{mle}$. The linear logistic regression model has been used for mortality prediction for similar data by many other authors, for example, Lemeshow et al. (1988).

To evaluate the "performance" of an individual hospital with respect to mortality, the severity measures are collected from the medical records of all patients entering the hospital with a qualifying condition during the preceding year. The 30-day mortality outcome and severity characteristics of the $n$ patients at a hospital during the year are represented by $Y_1, \ldots, Y_n$ and $X_1, \ldots, X_n$. The (hypothetical) death rates for all patients in the nation with the same severity characteristics they have been treated at hospital $H$ are denoted by

$$P_H(Y_1 = 1 | X_1), \ldots, P_H(Y_n = 1 | X_n) = P_{H,1}, \ldots, P_{H,n}.$$

The actual death rates of all patients in the nation with the same severity characteristics are denoted by

$$P_N(Y_1 = 1 | X_1), \ldots, P_N(Y_n = 1 | X_n) = P_{N,1}, \ldots, P_{N,n}.$$
We assume throughout that each patient in the hospital has an independent mortality outcome conditional on their condition's severity. For the medical conditions under consideration, this is a defensible assumption, since the conditions we study are not contagious. We also assume that the assignment of patients to hospitals may depend on the measured patient characteristics, $X$, but does not depend on any other unmeasured patient characteristics that are related to $Y$ after conditioning on $X$. This assumption means that patients with severity characteristics, $X$, at a hospital are regarded as equivalent to a random selection of all patients with the same severity characteristics. All estimates of hospital mortality rates must be judged with the knowledge that differences in the estimated hospital mortality rates may be due in part or in whole to remaining unmeasured differences in severity.

The observed death rate among the patients admitted to a hospital,

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i,$$

can be compared to the average of the estimated national death rates for patients with the same severity

$$\bar{P}_{N_{mle}} = \frac{1}{n} \sum_{i=1}^{n} \hat{P}_{N_{mle},i},$$

where the subscript $mle$ refers to the fact that the estimates are based on maximum likelihood estimates of the death rate from the MMPS data,

$$\hat{P}_{N_{mle},i} = \text{logit}^{-1}(\hat{\alpha}_{mle} + \hat{\beta}_{mle}X_i).$$

Although there is little contention about the comparison to be made, there are many competing statistical conceptualizations for evaluating the uncertainty in the difference between the adjusted national death rate and a particular hospital death rate. For the evaluations in this Note, we use indirect standardization. That is, the difference in these two death rates is viewed as an estimate of the difference between the (hypothetical) death rate if all patients in the nation with the same severity characteristics as the $n$ patients actually treated at the hospital were treated at that hospital, and the expected death rate of patients treated at hospitals across the nation with the same severity characteristics as the $n$ patients at the hospital,
\[ \bar{P}_H - \bar{P}_N = \frac{1}{n} \sum_{i=1}^{n} P_{H,i} - \frac{1}{n} \sum_{i=1}^{n} P_{N,i} \]
\[ = \frac{1}{n} \sum_{i=1}^{n} P_H(Y_i = 1|X_i) - \frac{1}{n} \sum_{i=1}^{n} P_N(Y_i = 1|X_i) . \] (3.2)

This difference in population rates is a function of the severity characteristics of the patients actually admitted to the hospital during the target period. It does not necessarily reflect the experience of any other patients who might be treated at the hospital in the future. All variances calculated in this Note are conditional on the characteristics of the patients actually observed at the hospital, \(X_1, \ldots, X_n\), and thus do not reflect any uncertainty about how the hospital rate and the national rates might differ on some other set of patients.

**Extrabinomial Variation**

To quantify the variation in hospital-specific death rates that exceeds the variation expected as a consequence of small sample binomial fluctuations, let \(n_i\) be the number of patients at the \(i^{th}\) hospital, \(\bar{Y}_i\) be the observed death rate at the hospital, and \(P_i\) be the underlying death rate at the hospital for any one of the medical conditions sampled by the MMPS. Under the assumption that patients are assigned to hospitals independent of their actual death probabilities, \(P_i\) is the hypothetical death rate that would be computed at the hospital if every patient in the nation with the medical condition could be observed being treated at the hospital. (This is a strong assumption that is difficult to verify.)

A model that can be used to explain extrabinomial variation observed assumes that the number of deaths at a hospital is binomial conditional on the death rate, \(P_i\), at the hospital, and conditional on the number of patients,

\[ n_i \bar{Y}_i \sim \text{Bin}(n_i, P_i) . \] (3.3)

The \(P_i\) however, are assumed to vary among hospitals according to a logit-normal distribution,

\[ \text{logit}(P_i) \sim N(\mu, \sigma^2_8) . \] (3.4)

Implicit in this formulation is the fact that \(n_i\) carries no information about \(P_i\); that is, \(n_i\) is ancillary.

Maximum likelihood estimates of the parameter \(\sigma_8\) in Eq. (3.4) are given in Section 4. The parameter \(\sigma_8\) measures the heterogeneity in the underlying hospital death rates, \(P_i\). The logit-normal formulation given in this Note is very similar to the beta-binomial distribution used by Jencks et al. (1988).
Incorporating Severity

A model that permits underlying hospital death rates to vary as a function of severity can be obtained by a simple expansion of the model in Eq. (3.1). Because patients from many different hospitals are represented, an additional subscript must be included to uniquely identify patients. Let the first subscript, $i$, indicate the hospital, and the second subscript, $j$, indicate the patient treated at the $i$th hospital. The expanded model from Eq. (3.4) including patient severity is

$$\text{logit}(P(Y_{ij} = 1|X_{ij})) = \alpha + \beta X_{ij} + \delta_i,$$

where $\delta_i$ represents how much the death rates (on the logit scale) for patients at hospital $i$ differ from the national rates for patients with the same $X$ values. The $\delta_i$ are assumed to vary among hospitals according to the logit-normal model

$$\delta_i \sim N(0, \sigma^2).$$

Maximum likelihood estimates of the variance component, $\sigma^2$, in models (3.5) and (3.6) are also given in Section 4. Computational details of the maximum likelihood estimators are given in Appendix F. In the context of models (3.5) and (3.6), $\sigma^2$ measures the heterogeneity in underlying hospital death rates after accounting for measurable differences in severity. How much heterogeneity exists in the underlying hospital-specific death rates is a key policy question, as is how much of this heterogeneity can be explained by differences in patient severity.

The model given by Eqs. (3.5) and (3.6) has been used by several authors to represent extrabinomial variability. Longford (1991) gives extensive references to applications of models of this type. More elaborate formulations that allow the hospital differences to vary depending on the values of $X$ and more general forms for the distribution of $\delta_i$ are not developed in this Note.

The logit-normal model in Eqs. (3.5) and (3.6) is a special case of the model in (3.3) and (3.4) with no covariates. These models form a nested sequence of models measuring the reduction in residual variation in underlying hospital death rates after adjustment for measured severity. The residual variance as measured by $\sigma^2$ must be interpreted carefully, however, because $\sigma^2$ does not necessarily decrease when additional predictor variables are added.
BAYES SHRINKAGE ESTIMATORS

We describe here our estimators of individual hospital mortality rates. We use Bayesian ideas from the variance component models presented in the previous section to calculate estimators based on the posterior distribution of \( \delta_i \). The shrinkage produced by these estimators accounts for the sampling fluctuations in the observed hospital mortality rates by moving the observed rates toward their values predicted by the patient severity characteristics. As has been shown in other contexts, such shrinkage generally results in improved estimates (Morris, 1983).

Consider the model for 30-day survival of patients at a specific hospital that raises or lowers the log odds of survival by the same amount, \( \delta \), for each patient,

\[
\text{logit}(P_H(Y_i = 1|X_i)) = \alpha + \beta X_i + \delta,
\]

where \((\alpha, \beta)\) are the same coefficients as in the national survival model. A prior distribution for \((\alpha, \beta, \delta)\) can be constructed in two stages. A prior distribution for \((\alpha, \beta)\) is formed using the maximum likelihood estimates, the information matrix, and the approximate sampling distribution from the MMPS data,

\[
f(\alpha, \beta) = N\left(\hat{\alpha}_{mle}, \hat{\beta}_{mle}, \hat{\delta}_{mle}\right).
\]

Because there is no compelling reason to change opinions about the quality of care or unmeasured severity at a particular hospital based on information in a national sample about the relative predictive strength of patient characteristics, we take the prior distribution of \( \delta \) to be independent of \((\alpha, \beta)\),

\[
\delta \perp \perp (\alpha, \beta).
\]

(3.9)

Assuming that the hospital effect, \( \delta_i \), has a normal prior distribution,

\[
\delta_i \sim N(0, \sigma_\delta^2)
\]

(3.10)
yields a logit-normal mixing model consistent with models (3.5) and (3.6).

Note that the prior or mixing distribution for \( \delta \) depends on the variance component, \( \sigma_\delta^2 \), which is also unknown but can be estimated from the PPS data as described in the preceding section. Thus, to obtain a prior distribution for \( \delta \), we must specify a prior distribution for \( \sigma_\delta^2 \) and then integrate to obtain

\[
P(\delta) = \int P(\delta|\sigma_\delta^2)P(\sigma_\delta^2)d\sigma_\delta^2.
\]
In the subsection "Constructing a prior for \( \delta \)" of Section 4, a scaled \( \chi^2 \) distribution is selected to describe the large uncertainty about \( \sigma^2_\delta \) that remains after it has been estimated using the national and PPS data. The choice of a scaled \( \chi^2 \) distribution for \( \sigma^2_\delta \) implies a scaled \( t_3 \) prior distribution for \( \delta \).

The prior distribution we have constructed is used for demonstrating the proposed Bayesian procedures in the simulation studies in Section 5. Because our information about the variance component \( \sigma^2_\delta \) is still very weak based on the estimates obtained from the PPS data, the prior distribution for \( \delta \) constructed in this Note is not recommended for actual application in the MMPS.

As \( \sigma_\delta \to \infty \), the Bayes procedure converges to the inferential procedures currently employed by the MMPS. Our choice of the scaled \( \chi^2 \) distribution for \( \sigma^2_\delta \) is conservative in the sense of erring on the side of larger values of \( \sigma^2_\delta \). This approach produces less shrinkage of \( \delta \) estimates toward zero and yields a procedure closer to the procedures currently used than a less "conservative" specification. On the other hand, the decision to err on the side of large \( \sigma^2_\delta \) is not "conservative" when considered from the viewpoint that a hospital should be labeled as exceptional only if there is strong evidence to support such a claim.

Appendix G gives a brief description of the numerical methods used to evaluate the point estimates (posterior means) and high posterior probability intervals.

A BAYESIAN DERIVATION OF THE ESTIMATOR OF HOSPITAL MORTALITY CURRENTLY USED BY THE MMPS

The estimator currently used by the MMPS is \( \overline{Y} - \overline{P}_{N\text{mle}} \) with a standard error given by

\[
\text{(sec)}^2 = \sum_{i=1}^{n} \hat{P}_{N\text{mle},i}(1 - \hat{P}_{N\text{mle},i}) + \text{var}\left(\overline{P}_{N\text{mle}}\right),
\]

where \( \text{var}\left(\overline{P}_{N\text{mle}}\right) \) is computed using the delta method applied to \( (\hat{a}_{\text{mle}}, \hat{\beta}_{\text{mle}}) \) in Eq. (B.1) of Appendix B.

Instead of using the normal mixing prior distribution in Eq. (3.10), consider the posterior distribution that results from the use of a flat, improper prior distribution for \( \delta \),

\[
f(\delta) \propto \text{const},
\]

which is a limiting case of the normal prior distribution as the prior variance for \( \delta \) becomes very large. The normal prior distribution in Eq. (3.8) with mean \( (\hat{a}_{\text{mle}}, \hat{\beta}_{\text{mle}}) \) and variance-covariance matrix \( \hat{\Sigma}_{\text{mle}} \) obtained from the large MMPS national sample is reused for \( (a, \beta) \).
The mode of the posterior distribution of \((\alpha, \beta, \delta)\) for the flat prior is denoted by \(\hat{\alpha}, \hat{\beta}, \hat{\delta}\), and the corresponding estimate of \(\bar{P}_H - \bar{P}_N\) is denoted by

\[
\bar{P}_H - \bar{P}_N = \frac{1}{n} \sum_{i=1}^{n} \log \left( \frac{\hat{\alpha} + \hat{\beta} X_i + \hat{\delta}}{1 + \hat{\beta} X_i} \right) - \frac{1}{n} \sum_{i=1}^{n} \log \left( \frac{\hat{\alpha} + \hat{\beta} X_i}{1 + \hat{\beta} X_i} \right)
\]

\[
= \frac{1}{n} \sum_{i=1}^{n} \hat{P}_{H,i} - \frac{1}{n} \sum_{i=1}^{n} \hat{P}_{N,i} .
\]

It is important to distinguish between \(\hat{\alpha}, \hat{\beta}, \hat{\delta}\), which are based on the new data at a specific hospital, and \((\hat{\alpha}_mle, \hat{\beta}_mle, \hat{\delta}_mle)\), which are based on the previously analyzed data from the MMPS national sample. The hybrid Bayes-maximum likelihood estimator \(\bar{P}_H - \bar{P}_N\) is a very close approximation to the estimator \(\bar{F} - \bar{F}_{N_mle}\) currently used in the MMPS and other hospital-level mortality reporting systems. The estimators differ slightly because the Bayes-based estimator employs \(\hat{\alpha}, \hat{\beta}\) instead of \((\hat{\alpha}_mle, \hat{\beta}_mle)\), but these estimates of \((\alpha, \beta)\) differ only slightly in the mortality applications because the prior distribution, which represents the information about \((\alpha, \beta)\) obtained from the large MMPS national sample, is much sharper than the information about \((\alpha, \beta)\) contained in the data from a single hospital. The near identity between \(\bar{P}_H - \bar{P}_N\) and \(\bar{F} - \bar{F}_{N_mle}\) is demonstrated with simulation results in Appendix C, and stated in concise mathematical form in the following theorem. A proof of this result is given in Appendix A, Eq. (A.5).

**Result:** Consider the estimator \(\bar{P}_H - \bar{P}_N\) computed using the mode \(\hat{\alpha}, \hat{\beta}, \hat{\delta}\) of the posterior distribution formed with the normal prior distribution for \((\alpha, \beta)\) in Eq. (3.8) and the flat prior distribution for \(\delta\) in Eq. (3.11). Suppose that the prior variance of \((\alpha, \beta)\) approaches 0; that is, \(\left\| \hat{K}_{mle} \right\| \rightarrow 0\). Then

\[
\bar{P}_H - \bar{P}_N \rightarrow \bar{F} - \bar{F}_{N_mle} .
\]

After the posterior mode of \((\alpha, \beta, \delta)\) has been computed, the asymptotic variance matrix for \((\alpha, \beta, \delta)\), denoted by \(\hat{K}\), can be computed by inverting the matrix of second derivatives of the log posterior distribution. A standard error for the estimated difference in the hospital and national survival rates, denoted by \(se_{mle}\), is computed using the the delta method in a fashion very similar to the one currently used by the MMPS. The derivation of \(se_{mle}\) is given in Appendix A. If \((\alpha, \beta)\) are taken to be exactly \((\hat{\alpha}_mle, \hat{\beta}_mle)\), only one estimating equation has to be solved for \(\hat{\delta}\), and the corresponding standard error calculations also simplify. This approach was not pursued, however, because it incorrectly underestimates the standard error by failing to account for the uncertainty in \((\alpha, \beta)\) that remains after estimation from the
large MMPS national sample. Figure C.1 in Appendix C shows that this underestimation of
the standard error can be as large as 20 percent in some realistic settings.

Although \( \text{se}_{\text{mle}} \) requires some additional calculations in order to find the mode,
\( (\hat{\alpha}, \hat{\beta}, \hat{\delta}) \), and the asymptotic covariance matrix, \( \hat{\Sigma} \), the overall computational effort is of a
similar magnitude to the current MMPS estimate because each method requires an extensive
double summation to compute estimates for \( \text{var}(\tilde{P}_N) \) from an approximating normal
distribution for \( (\hat{\alpha}, \hat{\beta}) \) as part of the calculation of the standard error for the difference in
death rates.

One difficulty with the Bayesian derivation using the flat prior distribution for \( \delta \) is
that a mode for \( \delta \) does not exist if all patients at a hospital survive. A small modification of
the estimator proposed in this section that corrects this deficiency by imputing a single
"observation" is also presented in Appendix A. Appendix B describes the standard error
currently used by the MMPS, and describes two additional competing alternatives. Appendix
C presents the repeated sampling coverage probabilities of the confidence intervals formed
using the standard errors based on a diffuse prior distribution for \( \delta \), and compares them to
the performance of intervals based on the other standard errors. The simulation study
showed that the confidence intervals based on \( \text{se_c} \) and \( \text{se}_{\text{mle}} \) have good coverage properties
for most settings likely to occur in practice, with \( \text{se}_{\text{mle}} \) offering a small improvement.

Once the estimator \( Y - \tilde{P}_{N,\text{mle}} \) has been embedded in a fully model-based framework,
alternative estimators and alternative estimands suggest themselves. One particularly
interesting estimand is hospital death rates standardized to a common population to make
estimates for different hospitals directly comparable. This is easily accomplished within the
model-based framework by selecting a mix of patient severity characteristics. The most
obvious choice, which we do not pursue here, is to select from a large random sample of the
national patient population. The estimate of \( \delta \) could also be used to form predictive values
for future patients.
4. RESULTS

OBSERVED DEATH RATES WITHOUT SEVERITY ADJUSTMENT

Hospital-specific annual death rates across the nation varied too much to be consistent with the null hypothesis of no differences in underlying death probabilities among the hospitals. The histogram in Figure 4.1 (a) presents the observed death rate for pneumonia for each eligible hospital in the United States based on the national database for fiscal year 1986. Figure 4.1 (b) simulates the histogram that would be observed if the null hypothesis of no difference between hospitals' underlying death rates was correct. Figure 4.1 (c) shows the differences between these histograms. There are too few hospital deaths near the middle of the observed rate histogram, and too many occurring far from the overall average rate to be consistent with the null hypothesis of no difference.

ESTIMATES OF THE VARIANCE COMPONENTS WITHOUT SEVERITY ADJUSTMENT BASED ON THE MEDICARE NATIONAL MORTALITY DATA

The variance component estimates from the logit-normal fit to the national database without any severity information are given in Table 4.1. From the reported standard errors, it is clear that the sample sizes in the national database are sufficient to accurately estimate the variance component, $\sigma^2$, which measures extrabinomial variability before any conditioning on severity. The variance component estimates differ too much across the different medical conditions to be consistent with the hypothesis that all medical conditions have the same variance component. However, the narrow range of the variance component estimates for the four medical conditions provides stable reference points for comparing the variance components after severity adjustment. That is, we do not anticipate large changes in the variance components, since the covariates measuring severity are only modestly strong predictors of survival, and further, we anticipate only a moderate amount of patient selection among hospitals.

ESTIMATES OF THE VARIANCE COMPONENTS WITH SEVERITY ADJUSTMENT BASED ON THE PPS DATA

Table 4.2 displays the maximum likelihood estimates for the logit-normal model in Eq. (3.5) with likelihood function given in Eq. (F.1) fit to the PPS data. The regression parameter estimates are roughly consistent with the regression parameter estimates given in Daley et al. (1988) and Keeler et al. (1990), and hence are not presented. There is strong evidence that the severity characteristics have predictive value.
Figure 4.1—Histogram of the Death Rate for Pneumonia Observed at Each Hospital (a), Histogram of Simulated Hospital Death Rates Assuming No Differences Between Hospitals (b), and Difference Between These Two Histograms (c)
Table 4.1

Variance Component Estimates Without Severity Conditioning Based on the National Database for 1986

<table>
<thead>
<tr>
<th>Medical Condition</th>
<th>$\hat{\sigma}_S$</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stroke</td>
<td>0.260</td>
<td>0.008</td>
</tr>
<tr>
<td>Pneumonia</td>
<td>0.283</td>
<td>0.006</td>
</tr>
<tr>
<td>Myocardial infarction</td>
<td>0.209</td>
<td>0.007</td>
</tr>
<tr>
<td>Congestive heart failure</td>
<td>0.211</td>
<td>0.007</td>
</tr>
</tbody>
</table>

Table 4.2

Variance Component Estimates with Severity Adjustment Based on the PPS Data

<table>
<thead>
<tr>
<th>Medical Condition</th>
<th>1981–82 (Standard Error)</th>
<th>1985–86 (Standard Error)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stroke</td>
<td>0.50 (0.20)</td>
<td>0.00 (0.22)</td>
</tr>
<tr>
<td>Pneumonia</td>
<td>0.33 (0.40)</td>
<td>0.00 (0.44)</td>
</tr>
<tr>
<td>Myocardial infarction</td>
<td>0.39 (0.20)</td>
<td>0.07 (0.79)</td>
</tr>
<tr>
<td>Congestive heart failure</td>
<td>0.50 (0.29)</td>
<td>0.48 (0.23)</td>
</tr>
</tbody>
</table>

The estimates of the variance components vary widely and have large standard errors. Hence, the PPS data offer little insight into the “between hospital” variation in the death rates. Only the 1985-86 data for congestive heart failure rule out a value of zero for the variance component on statistical grounds ($p < .05$). Even for this condition, caution must be exercised in interpreting the significance probabilities because of multiple comparison issues, since the national mortality data and our knowledge of the treatment of the four medical conditions make the four conditions approximately exchangeable. Thus, the larger variance component estimates for congestive heart failure could be explained as the result of selecting the largest from four highly variable sets of estimates.

For the purposes of the PPS study, a stratified sample of hospitals was included that overrepresented hospitals treating a high proportion of Medicaid patients. As a consequence, we anticipate that the hospital mortality rates in the PPS sample will vary more than those of the population of hospitals, resulting in larger variance components (Hartz et al., 1989). Nonetheless, logit-normal model fits to the PPS data without the severity covariates yielded variance component estimates consistent with the estimates from the national mortality data. Moreover, model fits like those in Eq. (3.5), but with the stratifying variables included as regressors, did not greatly reduce the variance component estimates.
To describe the maximum likelihood estimates more completely, Figure 4.2 displays minus twice the log likelihood function plotted against $\sigma_5$, with the regression parameters fixed at their maximum likelihood estimates. The likelihood functions are very flat in the range from 0.0 to 0.6 for all eight condition/time period combinations, but most plots begin to turn up sharply for values of $\sigma_5$ larger than about 0.6. Values of minus twice the log likelihood function that are larger than four provide increasingly strong evidence against the corresponding values of $\sigma_5$.

COMBINING INFORMATION ABOUT VARIANCE COMPONENT ESTIMATES FROM DIFFERENT PPS DATA SOURCES

If we think of the medical conditions and time periods as being roughly exchangeable, an improved estimate can be obtained by a pooling of the variance component estimates across time periods and medical conditions. We combined the various conditions using simple averages rather than using more elaborate model-based pooling of the data. The estimates of $\sigma_5$ in Table 4.2 averaged across the two time periods for the four medical conditions—stroke, pneumonia, myocardial infarction, and congestive heart failure—are 0.25, 0.17, 0.23, and 0.49. The value of the estimates of the variance components averaged over all the medical conditions and time periods is 0.28, consistent with the variance component estimates from the national data fit without severity covariates, and also consistent with the variance component estimates based on the PPS data fit without severity covariates, which have an average value of 0.22.

Based on the plots and the discussion in the above, values for $\sigma_5$ larger than 0.6 are unlikely and values of $\sigma_5$ larger than 1.0 are extremely unlikely, but the relatively small sample sizes in the PPS data do not allow us to distinguish $\sigma_5$ values in the range of 0.0 to 0.6.

CONSTRUCTING A PRIOR FOR $\delta$

The analysis of the national database provides very accurate estimates of the amount of extrabinomial variation in hospital death rates as measured by the logit-normal model. But these estimates are of the variation excluding the covariates measuring severity. However, we do not anticipate large changes in $\sigma_5^2$ from adjustment, although changes in either direction for $\sigma_5^2$ are possible.

The PPS data provide direct, but weak, information about the extrabinomial variation remaining after severity adjustment. We can conclude only that $\sigma_5^2$ with severity adjustment included is unlikely to be above 0.60 for any of the medical conditions, and we cannot absolutely rule out a value of 0. It is apparent from the analysis of the PPS data that
Figure 4.2—Minus Twice the Log Likelihood Function for $\sigma_\delta$ Evaluated at the Maximum Likelihood Estimates of the Regression Parameters
we still lack information to make adequately informed judgments about the size of the variance in hospital performance after severity adjustment. Appendix D contains calculations about the size and composition of a sample that would provide adequate estimates of $\sigma^2_\delta$. To demonstrate the proposed Bayes shrinkage estimator, we now construct a prior distribution that reflects our vague information about $\sigma^2_\delta$. Although we use this prior distribution for demonstration purposes in our simulation studies, we do not recommend its application in the MMPS.

One device for combining the information from the national and PPS data sources is to think about the distribution of $\sigma^2_\delta$ that would be obtained from knowing an independent, identically distributed sample of $\delta$ values, which would yield a direct estimate,

$$\hat{\sigma}_\delta^2 = \frac{\sum_{i=1}^{n}(\delta_i - \bar{\delta})^2}{(n - 1)}.$$

The information from our datasets should be represented by the information obtained from a very small direct sample of $\delta$ values. The uncertainty about $\sigma^2_\delta$ after adjustment from the PPS or MMPS data is large compared to the uncertainty about $\sigma^2_\delta$ before adjustment from the national data among the four medical conditions, and we have no substantive reason for thinking that adjustment has a differing effect on $\sigma^2_\delta$ across the different medical conditions; thus we assume a common prior distribution. We make a “conservative” choice of $\hat{\sigma}_\delta = 0.30$ as our best estimate of $\sigma_\delta$ in the severity adjustment model, which is larger than any of the estimated variances computed without severity adjustment. When the uncertainty in $\sigma_\delta$ is expressed as if it came from a small, independent, identically distributed sample of $\delta$ values (with $\delta$ normally distributed), the prior for $\sigma^2_\delta$ has the form

$$\sigma^2_\delta - \hat{\sigma}_\delta^2 \sim \frac{n-1}{\chi^2_{n-1}}.$$

We explored several choices of $(n - 1)$ and chose $n - 1 = 3$ as being conservative. This value expresses the most uncertainty about $\sigma^2_\delta$ possible within this framework while still producing a prior distribution for $\sigma^2_\delta$ (and $\delta$) which has finite mean and variance. This prior distribution for $\sigma_\delta$ posits only an 8 percent chance that $\sigma_\delta < 0.20$, a 32 percent chance that $\sigma_\delta$ is between 0.20 and 0.30, a 24 percent chance that $\sigma_\delta$ is between 0.30 and 0.40, a 22 percent chance that $\sigma_\delta$ is between 0.40 and 0.60, and a 14 percent chance that $\sigma_\delta > 0.60$. From what we know, this choice of prior distribution gives very generous weight to large values of $\sigma_\delta$ at the expense of values of $\sigma_\delta < 0.20$. Using standard Bayesian calculations for normal models, like those in Box and Tiao (1973), it follows that the prior distribution for $\delta$ is $\delta = \hat{\sigma}_\delta \hat{\delta}_\delta$. 

5. SIMULATION STUDIES COMPARING THE SHRINKAGE ESTIMATOR AND THE ESTIMATOR CURRENTLY USED BY THE MMPS

We report here on the statistical properties of the estimators of severity-adjusted individual hospital mortality rates. Comparisons are made by applying the estimators to simulated hospital mortality data where the “truth” is known. The simulated data vary across medical condition, patient mix, hospital size, and underlying hospital mortality rate. Two estimators are compared: the Bayes posterior mean estimator and high coverage intervals presented in Sections 3 and 4, and the current MMPS estimator and its standard error $se_c$. Outcomes include confidence interval coverage, mean square error of estimator, bias, power of tests, and selection of outlying hospitals.

The simulation studies evaluate the estimators under the assumption that the hospital and national death rates are distributed according to the models given in Eqs. (3.7) and (3.1). The simulation studies cover only these idealized conditions. They do not evaluate robustness to the modelling assumptions. However, they do include parameter values at the outer limits of values likely to be encountered in practice. The difference in the hospital and national death rates given in Eq. (3.2) is used as the target estimand to evaluate each of the estimators.

Two simulation studies are reported in this section. One study generates a large population of hypothetical hospitals to simulate our best information about the current hospital-level mortality distribution and its sources. This study allows us to summarize the overall statistical properties of the proposed estimators. The second study simulates many replications from each of a small number of fixed hypothetical hospitals to evaluate how the estimation procedures perform on specific types of hospitals (i.e., patient severity mix, hospital size, and quality). This study includes some extreme hospital types to highlight trends in the behavior of the estimators.

OVERVIEW OF SIMULATION RESULTS

By construction, the Bayes-based procedure favors the hospital with small differences between it and the national norm, producing a large estimated difference only when data from the hospital provides strong evidence to support a large difference. As a consequence, using confidence interval coverage, power, and bias as measures of performance, for repeated samples drawn with fixed parameter values, we expect the Bayes-based procedure to perform
well for settings with small values of $|\delta|$, and we expect the performance to degrade for large values of $|\delta|$. These expectations were confirmed by the simulation studies which quantify these trends in the hospital-level mortality application.

The simulations show that both estimation procedures produce good confidence intervals when the coverage probabilities are taken over a large population of hospitals. When specific hospital types are considered, the Bayes shrinkage procedure tends to undercover hospitals with small quality differences, and undercover hospitals with large quality differences. The coverage of the confidence intervals for the current MMPS procedure is approximately the same (and good) regardless of the hospital type.

The Bayes shrinkage estimator has a very large advantage in accuracy and stability. It outperforms the current MMPS estimator by over 300 percent in typical populations as measured by the mean square error. The Bayes shrinkage estimator retains a large advantage over $\hat{Y} - \hat{P}_{N_{\text{mle}}}$ for all hospital types except those with the most extreme quality differences, and the shrinkage estimator performed at least as well in those settings. The Bayes shrinkage estimate obtains the large improvement in mean square error despite the fact that it is substantially biased toward smaller estimates of quality differences when it is applied to a hospital with a large quality difference. The current MMPS estimator shows no evidence of bias in the simulation studies.

The Bayes shrinkage estimation and the current MMPS estimation procedure produce similar rankings of the extreme hospitals, and neither has much power to detect extreme hospitals under likely scenarios. The equivalents of significance levels and confidence intervals based on the Bayesian procedure, however, have correct coverage properties when applied to samples of hospitals identified as extreme by their estimated quality. The Bayesian calculations retain this important property because they explicitly account for the numerous quality estimates being made through the mixing distributions. The current MMPS estimation procedure, in contrast, produces misleading significance levels and confidence intervals when applied to samples of hospitals selected by their extreme estimated quality values because the procedure fails to account for the large "multiple comparison" problem. Note that the poor coverage properties of the MMPS procedure for selected samples of hospitals occur despite the good coverage properties that the procedure has for unselected samples of hospitals. In practice, most attention is focused on hospitals and medical conditions identified by their extreme estimates, so coverage properties of the statistical procedures are most important in this situation.
DESIGN OF THE SIMULATION STUDY WITH REPEATED SAMPLING FROM A SMALL NUMBER OF HYPOTHETICAL HOSPITALS

The simulation is arranged in a factorial design with four factors:

- 4 medical conditions (stroke, pneumonia, acute myocardial infarction, and congestive heart failure),
- 3 hospital severity levels (low, medium, high),
- 3 hospital sizes (25, 50, 100),
- 5 hospital qualities \( (\delta = -0.60, -0.30, 0, 0.30, 0.60) \).

The four medical conditions in the simulation are the four conditions used in the MMPS. The overall death rates, the covariates entering the death rate models, and the role of these covariates in the death rate models vary among these conditions (Jencks et al., 1988). For each medical condition, hypothetical hospitals with differing levels of severity in their patient populations were constructed using data from the MMPS national sample. Details of the construction of the hypothetical hospitals are given in Appendix I.

Survival data were simulated for patients at each hypothetical hospital using the model in Eq. (3.7). The underlying mortality rates at the hospital were simulated at five different levels. The levels (as measured by \( \delta \)) were selected to represent the outer limits of plausible differences in hospital mortality as determined in Section 4. The mortality outcome for the \( i^{th} \) patient at the hospital was generated from an independent Bernoulli distribution with probability \( P_{h,i} \), where the \( P_{h,i} \) were computed from the model in Eq. (3.7). The maximum likelihood estimates from the MMPS sample, \( (\hat{\alpha}_{mle}, \hat{\beta}_{mle}) \), were used as the population values, \( (\alpha, \beta) \), in the simulation.

For each setting in the factorial design, many simulated replications of the estimation procedures were produced. Each replication included the creation of a simulated MMPS national estimate, and simulated outcomes for each patient at the hypothetical hospital. The simulated MLEs from the MMPS national sample are denoted by \( (\hat{\alpha}_{sim}, \hat{\beta}_{sim}) \). Details of the generation of \( (\hat{\alpha}_{sim}, \hat{\beta}_{sim}) \) are given in Appendix H. Each simulation setting was replicated \( n_{sim} = 100 \) times. The severity information in the MMPS sample, \( X^m_1, \ldots, X^m_n \), and the severity information for the \( n \) patients at the hypothetical hospital are the same for each simulation replication.
DESIGN OF THE SIMULATION STUDY WITH REPEATED SAMPLING FROM FIXED HOSPITAL SETTINGS

Additional simulations were performed that retained much of the previous simulation design, but randomly selected the underlying hospital mortality index, represented by \( \delta \), from distributions chosen to represent the likely ranges of \( \delta \) based on the data analysis presented in Section 4. The additional simulations were repeated for \( \delta \) distributions representing low, medium, and high variation in underlying hospital mortality. A normal distribution with mean zero was used to represent each distribution with known standard deviations of 0.15, 0.25, and 0.50. By randomly selecting the \( \delta \) values, the operating characteristics of the Bayes-based procedure and current MMPS procedure can be evaluated under conditions more typical of those encountered in practice than the settings selected for the fixed hospital simulations.

The size of the sample at each hospital was also randomly selected (independent of \( \delta \)) according to the distribution \( n = 25, 50, 100, \text{ and } 200 \) with corresponding probabilities 0.35, 0.35, 0.25, 0.05, yielding an average hospital size of 61, which is in rough agreement with the hospital sizes in the national data summarized in Table 2.2 in Section 2. The high, medium, and low severity levels of patients at each hospital were also selected randomly (independent of \( \delta \) and hospital size) according to the probabilities 0.25, 0.50, and 0.25. The choice of distribution for the differences in severity levels was somewhat arbitrary, since we do not know this distribution. The distribution selected, along with the rather extreme nature of the simulated low- and high-severity hospitals, ensures that the patient severity levels at the simulated hospitals vary sharply, more than we think is likely in reality.

The hospital size and patient severity factors were randomly selected from empirically based distributions to give an indication of the operating characteristics of the statistical procedures for typical settings. The simulated populations of hospitals are also realistic enough to allow evaluation of how well the statistical procedures are able to identify hospitals with extreme values of \( \delta \).

The performance of the estimators did not appear to vary with the different medical conditions except through severity (marginal death rates). Therefore, only one medical condition, myocardial infarction, was simulated to reduce the computational burden. Two thousand replications (\( n_{sim} = 2000 \)) were performed to obtain reliable estimates of the characteristics of the hospitals identified as extreme cases. The large number of simulation replications also indicates the large number of hospitals available for extreme value screening.
Each replication of the simulation consists of the following steps:

1. Randomly select a value of δ, hospital size, n, and a severity level for the patient's condition,

2. Select a sample without replacement of n patients from a pool of 200 patients with the appropriate severity mix (100 patients were used in the pool in the first simulation) to obtain severity information for each simulated patient,

3. Generate a 0/1 mortality outcome for each patient according to the model in Eq. (3.7) using the MLE values of (α, β) from the MMPS study, and the randomly selected value of δ,

4. Simulate MLE estimates from the MMPS study as in the previous simulation.

SIMULATION ESTIMATES OF THE STATISTICAL PROPERTIES OF THE ESTIMATORS

Coverage probabilities were estimated for the nominal 95 percent confidence intervals computed using the current MMPS estimator $\bar{Y} - \bar{P}_{N_{mle}}$ and the current MMPS standard error se_c. Coverage probabilities were also computed using the "true" standard error based on parameter values used to create the simulation data, denoted by se_pop, for use as a reference indicating optimal performance achievable within this framework. The coverage probabilities are estimated from the simulation data by forming the confidence intervals and computing the proportion of intervals that include the population values used to simulate the data.

The power of the test of the hypothesis that $\bar{P}_H - \bar{P}_N = 0$ based $\bar{Y} - \bar{P}_{N_{mle}}$ was estimated in a similar fashion by computing the proportion of the simulation replications with $z = \left( \frac{\bar{Y} - \bar{P}_{N_{mle}}}{se_c} \right) < -1.96$ and $z = \left( \frac{\bar{Y} - \bar{P}_{N_{mle}}}{se_c} \right) > 1.96$.

Confidence intervals based on the posterior distribution from the Bayes-based procedures were formed by finding (lb, ub) such that under the Bayesian model, $P(\bar{P}_H - \bar{P}_N < lb) = 0.025$ and $P(\bar{P}_H - \bar{P}_N > ub) = 0.025$. The Bayes posterior mean, $\bar{P}_{H_b} - \bar{P}_{N_b} = E(\bar{P}_H - \bar{P}_N)$, was computed for use as a point estimate to compare to the standard MMPS estimate, $\bar{Y} - \bar{P}_{N_{mle}}$. Analogous to a power calculation, an estimate was judged to be different from zero if the Bayes probability satisfied $P(\bar{P}_H - \bar{P}_N > 0) < 0.025$ or $P(\bar{P}_H - \bar{P}_N < 0) < 0.025$.

The bias and mean square error (MSE) of the estimator $\bar{Y} - \bar{P}_{N_{mle}}$ were estimated from the simulation data by
\[ \text{BIAS} = \frac{1}{n_{\text{sim}}} \sum \left( \overline{Y} - \overline{F}_{N_{\text{mle}}} - (\overline{F}_H - \overline{F}_N) \right) \]

\[ \text{MSE} = \frac{1}{n_{\text{sim}}} \sum \left( \overline{Y} - \overline{F}_{N_{\text{mle}}} - (\overline{F}_H - \overline{F}_N) \right)^2 . \]

Analogous estimates were formed for the Bayes estimator \( \overline{F}_{H_b} - \overline{F}_{N_b} \). The mean absolute deviation of each estimator was also computed, but since it is well approximated by the square root of the MSE, the mean absolute deviation is not reported.

**Coverage Probabilities**

Results for the study simulating a population of hospitals are presented first. Table 5.1 displays the estimated coverage probabilities for the Bayes-based procedure and the MMPS procedure. Both procedures provide coverage probabilities in generally good agreement with the nominal 95 percent level, although the Bayes-based procedure begins to display undercoverage for the most extreme choice of the spread of the \( \delta \) distribution. The Bayes-based procedure achieves its good coverage properties by including more than 95 percent of the smaller \( \delta \) values within its 95 percent intervals, while including less than 95 percent of the larger \( \delta \) values within its 95 percent intervals. The coverage probabilities of the current MMPS procedure are very good and do not depend on the distribution of \( \delta \).

Coverage results for the fixed hospital simulation study are given in Table 5.2. The estimated coverage probabilities for the nominal 95 percent intervals for each level of \( \delta \) are averaged across all of the other simulation settings. The results for the Bayes-based procedure are presented first.

Among the four factors varied in the simulation, only \( \delta \) affected the coverage probability. The lower coverage probabilities for \( \delta = 0.60 \) reflect the fact that the likelihood function is skewed with a longer left tail. The range of estimated coverage probabilities for settings with \( \delta = 0.60 \) is from 0.61 to 0.82, consistent with the simulation error of \( \sqrt{0.7 \times 0.3 / 100} \approx 0.05 \).

The asymmetry in the posterior distribution resulting from the asymmetry in the

| Table 5.1 |
|---|---|---|
| Coverage Probabilities When \( \delta \) Are Randomly Generated |
| Estimator | Standard Deviation of the \( \delta \) Distribution |
| | 0.15 | 0.25 | 0.50 |
| Bayes | 0.98 | 0.97 | 0.91 |
| \( \overline{Y} - \overline{P}_{N_{\text{mle}}} \) | 0.95 | 0.95 | 0.95 |
Table 5.2
Coverage Probabilities for Nominal 95 Percent Bayes-Based Intervals for Each \( \delta \) Value Averaged Across the Simulation Settings

<table>
<thead>
<tr>
<th></th>
<th>0.6</th>
<th>0.3</th>
<th>0.0</th>
<th>0.3</th>
<th>0.6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bayes</td>
<td>0.93</td>
<td>0.99</td>
<td>0.99</td>
<td>0.92</td>
<td>0.73</td>
</tr>
<tr>
<td>MMPS</td>
<td>0.97</td>
<td>0.96</td>
<td>0.95</td>
<td>0.93</td>
<td>0.92</td>
</tr>
<tr>
<td>MMPS with population SE</td>
<td>0.95</td>
<td>0.95</td>
<td>0.95</td>
<td>0.95</td>
<td>0.95</td>
</tr>
</tbody>
</table>

likelihood function is apparent in the markedly different estimated coverage probabilities for \( \delta = -0.60 \) and \( \delta = 0.60 \), which is a consequence of the fact that most patients have low probability of death. Thus, even if all patients at a hospital survive, it is still unclear whether a hospital's underlying mortality is zero or merely very low. A hospital with a high underlying mortality rate, however, can be identified much more accurately because many more patients at such a hospital would die than expected from the national experience. The resulting uncertainty about the potential for a hospital to have exceptionally low underlying mortality causes confidence intervals to cover the lower values of \( \delta \) even when the posterior mean and mode are close to zero.

The intervals currently reported by the MMPS,

\[
\left(-1.96 \times se_c + \left( \bar{Y} - \bar{P}_{N,\text{mle}} \right), 1.96 \times se_c + \left( \bar{Y} - \bar{P}_{N,\text{mle}} \right)\right),
\]

have generally good coverage properties, summarized in Table 5.2. The corresponding coverage probability estimates based on intervals computed using the population values and the correct formula for the standard errors are also given in Table 5.2. Additional details for \( se_c \) and comparison to other standard errors are given in Appendix C.

Although \( se_c \) from the MMPS produced confidence intervals with generally good coverage properties, more careful examination showed that it varied substantially depending on the patient severity and relative mortality parameter (\( \delta \)) of the hospital. Figure 5.1 shows the estimated coverage probabilities for each simulation setting computed using the MMPS (\( se_c \)) plotted against the national death rate for patients like those at the hospital. The points marked with an \( \times \) have \( \delta = -0.60 \); those marked with a \( \circ \) have \( \delta = 0.60 \). The respective overcoverage and undercoverage for these two values of \( \delta \) results because the correct variance formula, \( P_{H,i}(1-P_{H,i}) \), increases with \( \delta \) and corresponds to the variance
used in the MMPS ($se_c$) only when $\delta = 0.0$. (The two points in the upper left corner marked with a plain dot correspond to $\delta = -0.30$.) Note that this over- and undercoverage pattern decreases as $\overline{P}_N$ nears 0.5, but it is still very apparent (though less severe) for $\overline{P}_N = 0.2$. The degradation in the performance occurs when $\delta = \pm 0.60$, which is at the outer limits of plausible values for $\delta$. The performance of $se_c$ is much better for intermediate values of $\delta$.

The coverage probabilities do not appear to vary with the hospital sample size. In particular, the plot in Figure 5.1 would be similar for much larger hospital sample sizes. The MMPS intervals based on $se_c$ are an example of a procedure that is not asymptotically correct, but is approximately correct over a large variety of the settings that are most likely to be encountered in practice.
Mean Square Error of Estimation

The Bayes-based procedure, even with the relatively diffuse prior being employed, is much more conservative in judging that there is evidence to identify a poor hospital. There is little indication from the coverage estimates to support the use of the Bayes-based procedure over the existing alternatives. The Bayes-based procedure, however, provides much more accurate estimates of $\bar{P}_H - \bar{P}_N$. Estimates of the MSE of $\bar{y} - \bar{P}_{N_{MLE}}$ divided by the MSE of the Bayes posterior mean,

$$\frac{\mathbb{E}\left[\left(\bar{y} - \bar{P}_{N_{MLE}}\right) - (\bar{P}_H - \bar{P}_N)\right]^2}{\mathbb{E}\left[(\bar{P}_{H_{bayes}} - \bar{P}_{N_{bayes}}) - (\bar{P}_H - \bar{P}_N)\right]^2},$$

from the study that simulates a population of hospitals are 3.96, 3.18, and 1.68 when the standard deviations of the $\delta$ distribution are 0.15, 0.25, and 0.50. This large advantage in the MSE for the Bayes shrinkage estimate applies across all values of $\delta$, although it is much smaller for large values of $\delta$.

Table 5.3 displays the estimates of the mean square error of $\bar{y} - \bar{P}_{N_{MLE}}$ divided by the MSE of $\bar{P}_{H_{bayes}} - \bar{P}_{N_{bayes}}$ for the simulation study with fixed hospital settings. Table 5.3 shows that the Bayes-based estimates are substantially better than the standard estimate, $\bar{y} - \bar{P}_{N_{MLE}}$, for $\delta$ values -0.30, 0.00, and 0.30, and retain their superiority for settings with $\delta = \pm 0.60$, except when $n = 100$, where they are about equal. The ratios of the mean square errors in Table 5.3 are averaged across medical conditions and severity settings. There is some variation in MSE associated with different severity settings. The likelihood functions for data from higher severity settings have more information about hospital differences so that there is less gain from the use of the prior distribution in settings with higher severity.

The Bayes-based estimates achieve their superiority by producing much more stable estimates of hospital differences that are closer to zero. The improved performance of the Bayes-based procedure occurs in spite of the fact that $\bar{P}_{H_{bayes}} - \bar{P}_{N_{bayes}}$ is a biased estimate for large values of $|\delta|$. Table 5.4 displays the relative bias, $\left|\frac{\mathbb{E}(\bar{P}_{H_{bayes}} - \bar{P}_{N_{bayes}}) - (\bar{P}_H - \bar{P}_N)}{\bar{P}_H - \bar{P}_N}\right|$, averaged across medical conditions and severity settings. The Bayes-based procedure shrinks large estimates of hospital differences very strongly toward zero, but because of the increased stability, the Bayes-based procedure provides better estimates than $\bar{y} - \bar{P}_{N_{MLE}}$, for which there is no evidence of bias, but which are extremely variable.

A Bayes shrinkage model based on normally distributed observations would be expected to have constant relative bias across all $\delta$ values. This is approximately true for the binary data shrinkage model, but there is some difference in the relative bias between
Table 5.3

<table>
<thead>
<tr>
<th>Values of $\delta$</th>
<th>$n$</th>
<th>25</th>
<th>50</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>2.10</td>
<td>1.24</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>0.30</td>
<td>6.07</td>
<td>2.96</td>
<td>2.09</td>
<td></td>
</tr>
<tr>
<td>0.00</td>
<td>10.32</td>
<td>5.82</td>
<td>3.85</td>
<td></td>
</tr>
<tr>
<td>-0.30</td>
<td>4.51</td>
<td>2.86</td>
<td>2.19</td>
<td></td>
</tr>
<tr>
<td>-0.60</td>
<td>1.84</td>
<td>1.19</td>
<td>1.02</td>
<td></td>
</tr>
</tbody>
</table>

Table 5.4

<table>
<thead>
<tr>
<th>Relative Bias, $\frac{</th>
<th>E(\bar{P}<em>{Hb} - \bar{P}</em>{N_{mle}}) - (\bar{P}_H - \bar{P}_N)</th>
<th>}{</th>
<th>\bar{P}_H - \bar{P}_N</th>
<th>}$, of the Bayes Procedure Averaged Across Medical Conditions and Severity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Values of $\delta$</td>
<td>$n$</td>
<td>25</td>
<td>50</td>
<td>100</td>
</tr>
<tr>
<td>-0.60</td>
<td>0.69</td>
<td>0.52</td>
<td>0.38</td>
<td></td>
</tr>
<tr>
<td>-0.30</td>
<td>0.73</td>
<td>0.55</td>
<td>0.41</td>
<td></td>
</tr>
<tr>
<td>0.30</td>
<td>0.59</td>
<td>0.48</td>
<td>0.39</td>
<td></td>
</tr>
<tr>
<td>0.60</td>
<td>0.62</td>
<td>0.48</td>
<td>0.35</td>
<td></td>
</tr>
</tbody>
</table>

settings with positive and negative $\delta$ values because of the asymmetry that is present in the proportion scale.

In the study simulating a population of hospitals, the Bayes posterior mean and $\bar{Y} - \bar{P}_{N_{mle}}$ are unbiased; the unbiasedness of the Bayes posterior mean is due to the choice of symmetric distributions for $\delta$ in the simulation and would not necessarily obtain when asymmetric distributions are used for $\delta$.

Power of Tests

Table 5.5 gives power estimates for tests of no difference between the hospital and the nation at the 5 percent significance level for the simulation study with fixed hospital settings. Power is estimated from the proportion of simulated datasets yielding a significant two-sided test statistic for the alternative values of $\delta$ given in the first row of Table 5.5. The
Table 5.5

Estimates of Power for Testing the Null Hypothesis $P_H - P_N = 0$ at .05 Level

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Estimator</th>
<th>Alternative</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$\delta = -0.6$</td>
</tr>
<tr>
<td>n = 25</td>
<td>Bayes</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>$\bar{Y} - \bar{P}_{N,mle}$</td>
<td>0.12</td>
</tr>
<tr>
<td>n = 50</td>
<td>Bayes</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>$\bar{Y} - \bar{P}_{N,mle}$</td>
<td>0.21</td>
</tr>
<tr>
<td>n = 100</td>
<td>Bayes</td>
<td>0.37</td>
</tr>
<tr>
<td></td>
<td>$\bar{Y} - \bar{P}_{N,mle}$</td>
<td>0.44</td>
</tr>
</tbody>
</table>

Power estimates for the Bayes-based procedure averaged across severity levels and medical conditions are given in the upper entries of Table 5.5, and the power estimates for the current MMPS estimator and standard error are given in lower entries. The power of both procedures is very low for the more commonly occurring samples sizes and $\delta$ values. Even with the relatively diffuse prior distribution being employed in the Bayes procedure, we would seldom expect a hospital with 100 or fewer patients to provide enough evidence to rule out the possibility that the hospital performs as well as the national norm with a 5 percent error rate.

The power for the Bayes-based procedure is somewhat less than the power of the test based on the MMPS for negative $\delta$ values, but is much smaller for positive $\delta$ values. This is a consequence of the skewness in the left-hand tail of the posterior distribution.

Unlike the Bayes-based procedure, $\bar{Y} - \bar{P}_{N,mle}$ and $se_e$, which are currently used by the MMPS, have more power to detect low-quality hospitals than high-quality hospitals. This occurs because the current MMPS procedure forces symmetric confidence intervals, and the standard errors are smaller (on the log-odds scale) when the marginal death rates are higher, which occurs when $\delta \gg 0$.

The power of each procedure is also very low in the study that simulates a large collection of hospitals. The power of each procedure is estimated by the proportion of hospitals whose confidence interval does not include zero. When the standard deviations of the $\delta$ distribution are 0.15, 0.25, and 0.50, the power of the Bayes-based procedure is only 0.01, 0.03, and 0.12, and for the current MMPS procedure it is 0.07, 0.11, and 0.25.
Identifying Exceptional Hospitals

One possible use of the hospital-specific mortality data is to rank hospitals and attempt to identify exceptional hospitals. A simple ranking system was implemented using the Bayes-based procedures and the current MMPS procedure as part of the simulation study that generated a population of hospitals. Hospitals were ranked on the Bayes posterior probability that the hospital has mortality higher than the national rate, $P(\bar{P}_H - \bar{P}_N > 0)$, and each hospital was ranked based on the MMPS Z-statistic, $z = \left( \bar{y} - \bar{P}_{N,mle} \right) / se_c$. For each of these statistics, the 5 percent of the 2000 simulated hospitals with the highest values were identified as potentially “bad hospitals,” and the 5 percent with the lowest values were identified as potentially “good hospitals.” Five percent was used because the resulting samples contain 50 hospitals each, yielding stable estimates of the characteristics of the extreme hospitals. The difference in the means of the actual values of $\bar{P}_H - \bar{P}_N$ among the hospitals identified as extreme is the primary criterion for comparing the performance of the different ranking methods. The difference in means is an appropriate criterion because it is approximately proportional to the number of potentially “preventable deaths.” The alternative criterion using mean differences weighted by the hospital sample sizes yielded similar results and is not reported. The results of the simulation are shown in Table 5.6. To show the best results theoretically obtainable, the average of the extreme 5 percent values of $\bar{P}_H - \bar{P}_N$ is also included in Table 5.6.

It is clear from the table that the Bayes-based ranking procedure and the ranking MMPS procedure perform similarly. Figure 5.2 displays a histogram of the $\bar{P}_H - \bar{P}_N$ values for the 50 hospitals identified as the poorest hospitals using the MMPS Z-statistic, and for comparison, a histogram of the 50 hospitals with the highest values of $\bar{P}_H - \bar{P}_N$. The hospitals were generated with the intermediate (and most likely) value of the dispersion of $\delta$.

Table 5.6

The Mean of $\bar{P}_H - \bar{P}_N$ for Hospitals Identified as Extremely Low or High Quality by the Different Ranking Criteria

<table>
<thead>
<tr>
<th></th>
<th>Standard Deviation of the $\delta$ Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>Low</td>
</tr>
<tr>
<td>MMPS</td>
<td>0.020</td>
</tr>
<tr>
<td>Bayes</td>
<td>0.023</td>
</tr>
<tr>
<td>$\bar{P}_H - \bar{P}_N$</td>
<td>0.060</td>
</tr>
</tbody>
</table>
Figure 5.2—Histogram of the 50 Most Extreme Values of $\bar{P}_H - \bar{P}_N$ from the Simulation Sample of 2000 (a) and Histogram of the 50 Values of $\bar{P}_H - \bar{P}_N$ Selected Using the Current MMPS Z-statistic (b)

From Figure 5.2, it is apparent that although some extremely poor hospitals are correctly identified, the selection procedure incorrectly identified two (4 percent) of the hospitals, which are actually better than the national norm, and 16 percent of the hospitals identified as very poor were actually very close to the national norm. The average p-value for the 50 hospitals identified as extremely poor based on the Z-statistic is 0.001. With p-values this small, even a sophisticated user of statistics is tempted to conclude that the evidence against these hospitals is exceedingly strong. The average value of the Bayes probabilities, $P(\bar{P}_H - \bar{P}_N > 0)$, for the 50 hospitals identified as extremely poor is 0.02, in much closer agreement with the actual rate of false positives (there was only one false positive using the Bayes-based ranking).

The misleading performance of the current MMPS procedure when applied to hospitals selected because of their unusual estimates applies to confidence intervals also. The nominal 95 percent confidence intervals from the current MMPS procedure contained the population
difference in only $24/50 = 48$ percent of the hospitals selected because of their high estimated mortality rates. The corresponding Bayesian intervals include $48/50 = 96$ percent of these population differences.

The Bayesian probabilities, like the Bayesian posterior mean, give a much more conservative and accurate assessment of the evidence provided by the hospital-specific mortality data than the current MMPS procedures. The Bayes procedure employed here is still likely to be too generous in the weighting of the hospital-specific data, because the prior distribution employed favors larger values of $\delta$ than justified by current information.

Several other statistics were used to rank the hospitals, including $\bar{Y} - \bar{P}_{N_{\text{mle}}}$ and $\bar{P}_{H_{b}} - \bar{P}_{N_{b}}$. Although the precise ranking varied with each ranking variable, the hospitals that are identified as extreme change very little. This result is somewhat surprising in light of the literature comparing ranking methods based on Bayesian estimates and non-Bayesian methods such as Laird and Louis (1989) that suggest substantial gains for the Bayesian estimates. This could result largely because although hospitals with 25 and 200 patients have substantially different standard errors, there are very few of the latter hospitals, and the standard errors of the remaining hospitals are only moderately different.
6. DISCUSSION

DO HOSPITALS DIFFER IN THEIR UNDERLYING MORTALITY RATES?

It is abundantly clear from national data that hospitals differ in their underlying death rates after accounting for sampling variation (Park et al., 1990; Jencks et al., 1988). Specifically, there is extrabinomial variation in hospital-level death rates. Our variance component models confirm this finding across the same four medical conditions of stroke, pneumonia, myocardial infarction, and congestive heart failure. The variance component, which measures extrabinomial variability in the logit-normal model, corresponds to an estimated standard deviation, $\sigma_5$, ranging from 0.21 to 0.28 across medical conditions with standard errors of the estimate of less than 0.01 (Table 4.1). These values do not take into account any differences in patient mix across hospitals.

Our logit-normal variance component model using the PPS data confirms earlier findings from the MMPS data (Jencks et al., 1988) that individual patient outcomes vary with severity of condition and is consistent with reported analyses of the PPS data (Kahn et al., 1990a; Rogers et al., 1990). Although the PPS design is superior to the MMPS design for estimating the severity-adjusted extrabinomial variation in hospital death rates, $\sigma_5$, the variation is still poorly estimated. The data suggest different values of $\sigma_5$ for the four medical conditions and two time periods. Assuming a difference, the most that can be said for the PPS data is that $\sigma_5$ is between 0 and 0.6 (Table 4.2, Figure 4.2). Indeed, only the two congestive heart failure variance components can be distinguished from zero by statistical significance tests (Figure 4.2).

How to combine the variance components of the national data with no severity adjustment with the PPS data with severity adjustment is unclear. On the one hand, that the covariates measuring severity are only modestly strong predictors of survival (Keeler et al., 1992) suggests that using the range of 0.2 to 0.3 of accurately estimated $\sigma_5$ from the national data (without severity adjustment) is sensible, since it is the middle of the 0.0 to 0.6 range of $\sigma_5$ suggested by the PPS data. On the other hand, since the value of $\sigma_5$ with severity adjustment can be either smaller or greater than the value of $\sigma_5$ without severity adjustment, the 0.2 to 0.3 range has no guarantee of being valid.

POOLING

If we pool the estimates of $\sigma_5$ across the two time periods and four medical conditions, we get 0.28 for the national data without severity adjustment and 0.22 from the PPS data.
with severity adjustment; these are consistent with each other. Such pooling requires us to 
think of the medical conditions and time periods as exchangeable, a dubious assumption. 
However, in the PPS data, the increased individual biases from pooling may well be balanced 
by improved accuracy.

THE NEED FOR AN IMPROVED NATIONAL SAMPLE

Improved estimates of $\sigma_5$ would serve three important policy goals.

First, if $\sigma_5$ is very close to zero, i.e., there is no evidence of hospital quality differences 
after severity adjustment, then the need for any mortality reporting system would not be 
demonstrated. To date, the evidence about even this basic question is not definitive. The 
strongest evidence for substantial quality differences after severity adjustment comes from 
comparisons of different types of hospitals such as those made by Hartz et al. (1989).

Second, even if $\sigma_5$ is moderate in size, suggesting a possible quality problem, the 
expense involved in collecting severity information at each hospital may not be justified 
because there may be too little information available from the relatively small samples at 
most individual hospitals to make better decisions. The calculations in Appendix E are 
illustrative of this important problem.

Suppose $\sigma_5 = 0.25$, which represents moderately large variation consistent with our 
current knowledge, and which is certainly large enough to cause concern if this variation 
could be attributed to quality differences. In a typical example (stroke, $\delta_{mle} = 0$, medium 
severity in Table E.1), calculations like those in Appendix E show that for a hospital with 50 
patients in a given medical condition, our uncertainty about the quality (measured on the 
log odds scale) would be expected to be reduced from 0.25 standard deviation units to 0.22 
standard deviation units, or a 12 percent improvement, as a result of collecting and 
analyzing the severity-adjusted mortality rate at the hospital. Improvements of this 
magnitude are hard to justify in light of the expense involved in collecting severity 
information regardless of whether a government agency or an individual hospital is bearing 
the expense.

Improvements may be obtained by pooling data from several years and medical 
conditions. Improved estimates of $\sigma_5^2$, and more extensive analyses of the type given in 
Appendix E, are strongly recommended before any policy encouraging the collection of 
severity information is pursued.

The third goal that would be achieved by an improved estimate of $\sigma_5^2$ is to provide a 
stronger empirical basis for specifying a prior or mixing distribution for Bayes shrinkage 
estimation, which produces improved estimation of individual hospital rates.
IMPLICATIONS FOR INDIVIDUAL HOSPITAL REPORTING

HCFA currently releases reported Medicare death rates and standard errors by hospital for 17 diagnostic categories from national data and also supplies hospitals through the MMPS with a way to calculate their own severity-adjusted death rates and corresponding standard errors. Our Bayesian derivation of the MMPS estimator and its standard error (and its modification in Appendix A) is based on a diffuse prior distribution for \( \delta \), the hospital effect. Contrasting this diffuse prior to our estimates of \( \sigma_5 \), the variance component of the severity-adjusted hospital effects gives an interpretable comparison of the MMPS estimator and the Bayesian shrinkage estimators of hospital death rates.

For all but the largest hospitals and medical conditions, the information contributed by the mortality data at a single hospital is small relative to the information available from national data sources, which inform us about the likely range of hospital mortality differences. In short, the choice of the prior or mixing distribution matters, and the current practice of using a flat mixing distribution (on the log odds scale) is a poor choice whether the mixing distribution is invoked explicitly or implicitly.

The simulation comparisons of the MMPS procedure and the Bayes-based procedure for hypothetical hospital populations reveal some interesting conclusions. As expected, the Bayes shrinkage estimator performs relatively well for hospitals with death rates near the national norm (small \( |\delta| \)) and relatively less well for more extreme values (large \( |\delta| \)). This pattern holds for coverage probabilities, mean absolute error of estimation, and mean square error of estimation. The Bayes estimator produces biased estimates of hospital differences with a lower variance than the MMPS estimator. By shrinking large estimates of hospital differences strongly toward zero, the increased stability of the Bayes estimator gives much better estimates for hospitals that are not demonstrably extreme.

An important use of these estimators is in identifying exceptional hospitals. The Bayes procedure is more conservative in assessing the evidence that a hospital is extreme than the MMPS estimator. Overall, the two procedures perform similarly in identifying the 5 percent best and worst hospitals and both do rather poorly. The Bayes-based procedure, however, gives correct significance levels and confidence intervals for the hospitals identified by their extreme estimates. The current MMPS estimation procedure, in contrast, gives incorrect misleading significance levels and confidence intervals for hospitals selected on the basis of their estimated mortality rates that exaggerate the evidence about the extreme nature of these hospitals.
Appendix A
STANDARD ERRORS BASED ON A DIFFUSE PRIOR DISTRIBUTION

This appendix describes a minor modification to the standard error (SE) estimator, $se_{mle}$, based on the flat prior distribution for $\delta$. The modification is required because when $\bar{Y}$ equals zero, the logistic regression estimating equations does not have a maximum, so the estimator for $\delta$ does not exist. An estimator is proposed that uses a very diffuse prior distribution giving slightly more weight to values of $\delta$ near zero. The estimator is essentially equal to $\bar{Y} - \hat{P}_{N_{mle}}$. Formulas for the model-based asymptotic standard error of $\bar{Y} - \hat{P}_{N_{mle}}$ are derived at the end of this appendix. This prior distribution resolves the nonconvergence problem by shrinking $\delta$ slightly toward zero.

A PRIOR DISTRIBUTION BASED ON ONE IMPUTED OBSERVATION

The prior distribution used here is very similar to a prior distribution proposed by Clogg et al. (1991) that involves the imputation of data into the likelihood function. The principal advantage of this approach is that it can be implemented using standard logistic regression software by modifying the data that are input.

To construct this prior distribution, begin with a binomial likelihood function,

$$ P_i^{Y_{i1}}(1-P_i)^{Y_{i0}}, $$

where $Y_{i1}$ corresponds to the number of observations with $Y = 1$, and $Y_{i0}$ corresponds to the number of observations with $Y = 0$. The prior is formed by splitting $m$ (typically one) observation across the $n$ observations in the hospital sample, where the relative number of deaths and survivals is determined by the estimated death rate model with $\delta = 0$:

$$ Y_{i1} = \frac{\hat{P}_{H_{mle,i}(0)}}{n} m, \quad Y_{i0} = \frac{1 - \hat{P}_{H_{mle,i}(0)}}{n} m, $$

where

$$ \text{logit}(\hat{P}_{H_{mle,i}(\delta)}) = \hat{\alpha}_{mle} + \beta_{mle}X + \delta, $$

The prior distribution for $(\alpha, \beta, \delta)$ is then given by

$$ f(\alpha, \beta) \prod_{i=1}^{n} (P_{H,i}(\alpha, \beta, \delta))^{Y_{i1}} (1 - P_{H,i}(\alpha, \beta, \delta))^{Y_{i0}} \quad (A.1) $$
where

$$\text{logit}(P_{H,i}(\alpha, \beta, \delta)) = \alpha + \beta X + \delta,$$

and $f(\alpha, \beta)$ is the prior for $(\alpha, \beta)$ given in Eq. (3.8). The posterior distribution of $(\alpha, \beta, \delta)$ is then

$$\log(f(\alpha, \beta, \delta) | Y_1, \ldots, Y_n, X_1, \ldots, X_n) \propto -\frac{1}{2} \left( \frac{\alpha - \hat{\alpha}_{mle}}{\hat{\sigma}_{mle}} \right)^2 - \frac{1}{2} \left( \frac{\beta - \hat{\beta}_{mle}}{\hat{\sigma}_{mle}} \right)^2 + \sum_{i=1}^{n} (Y_i + \hat{\delta}_{mle}) (\alpha + \beta X_i + \hat{\delta}_{mle}) - \sum_{i=1}^{n} \left( 1 + \frac{m}{n} \right) \log \left( 1 + e^{(\alpha + \beta X_i + \hat{\delta}_{mle})} \right).$$

(A.2)

This has the same form as the logistic likelihood given in Eq. (G.1) and can thus be evaluated using the same software by adding $Y_{i1}$ and $Y_{i0}$ to the input data.

The prior distribution in Eq. (A.1) depends on the size of the hospital sample and on the severity characteristics at the hospital. However, it does not depend on the mortality experience at the hospital, so it is a true prior in this sense. The function in Eq. (A.1) does not necessarily integrate to a finite number so it may not be a proper prior distribution.

Deciding how much information to impute with the prior distribution is a largely arbitrary choice. The SE evaluated in this Note use $m = 1$, but other choices are feasible. The standard error derived using $m = 1$ is denoted by $se_{b1}$ and is derived below. The amount of information about $\delta$ supplied by this prior distribution relative to the amount of information supplied by the hospital data decreases as the size of the hospital increases. The information supplied by the prior is generally very small compared to information supplied by the hospital, so that the resulting posterior distribution and its mode are also very similar.

The estimates of $(\alpha, \beta, \delta)$ obtained by maximizing Eq. (A.2) are denoted by $(\hat{\alpha}_{b1}, \hat{\beta}_{b1}, \hat{\delta}_{b1})$, and the corresponding estimates of $\overline{P}_H - \overline{P}_N$ are denoted by $\overline{P}_{Hb1} - \overline{P}_{Nb1}$ with a SE estimated by $se_{mle}$. The estimator $\overline{P}_{Hb1} - \overline{P}_{Nb1}$ is very similar to the estimator for $\overline{P}_H - \overline{P}_N$ based on the flat prior distribution for $\delta$, $\overline{Y} - \overline{P}_N$, so that $\overline{P}_{Hb1} - \overline{P}_{Nb1}$ and $se_{b1}$ can be used as an alternative to $\overline{Y} - \overline{P}_N$ and $se_{mle}$. The similarity of $\overline{Y} - \overline{P}_N$ and $\overline{P}_{Hb1} - \overline{P}_{Nb1}$, and the coverage properties of these estimators and standard errors, are explored using simulations in Appendix C.

To better understand the prior distribution specified in Eq. (A.1), we develop an approximation for the marginal distribution of $\delta$,.
$$f(\delta) = \int f(\alpha, \beta) \prod_{i=1}^{n} (P_{H,i}(\alpha, \beta, \delta))^Y_{1i} \left(1 - P_{H,i}(\alpha, \beta, \delta)\right)^{Y_{10}} d(\alpha, \beta). \tag{A.3}$$

Since $f(\alpha, \beta)$ is a very peaked normal density relative to the remaining terms in Eq. (A.3), a good approximation for Eq. (A.3) is given by

$$f(\delta) = \prod_{i=1}^{n} \left(\frac{P_{H,i}(\hat{\alpha}_{mle}, \hat{\beta}_{mle}, \delta)}{1 - P_{H,i}(\hat{\alpha}_{mle}, \hat{\beta}_{mle}, \delta)}\right)^{Y_{1i}} \left(1 - \hat{P}_{H_{mle},i}(\delta)\right)^{Y_{10}} = \prod_{i=1}^{n} \left(\hat{P}_{H_{mle},i}(\delta)\right)^{Y_{1i}} \left(1 - \hat{P}_{H_{mle},i}(\delta)\right)^{Y_{10}}. \tag{A.4}$$

Figure A.1 displays a plot of Eq. (A.4) normed so that the approximate density has a maximum value of one. The likelihood is evaluated for a simulated hospital of size 25 with severity distribution and mortality that is typical for pneumonia cases. Similar densities are obtained for different hospital sample sizes. The shape of the density can change with the severity of the patients' conditions in a hospital. The density is extremely diffuse; a logit value of five corresponds to implausibly strong hospital differences. The skewness results from the fact that the national death rate is about 20 percent. Because we imputed an observation consistent with this rate for the hospital, the prior eliminates large positive values for $\delta$ that would have almost certainly resulted in a death for the imputed patient. However, the observation imputed does not eliminate the possibility of very large negative values for $\delta$. It gives only enough preference to small $|\delta|$ values to ensure a mode for the posterior distribution in Eq. (A.2).

STANDARD ERRORS

The following gives a brief outline of the calculation of $se\left(\hat{P}_H - \hat{P}_N\right) = se_{mle}$. The same calculations are used for $se\left(\hat{P}_{H_{b1}} - \hat{P}_{N_{b1}}\right) = se_{b1}$ with small modifications to the input values.

The calculation of the standard error is performed from a Bayesian perspective which regards $(\alpha, \beta, \delta)$ as random variables, and $(\hat{\alpha}, \hat{\beta}, \hat{\delta})$ as fixed, observed quantities. Expressing $\hat{P}_H - \hat{P}_N$ as a function of $(\alpha, \beta, \delta)$,

$$f(\alpha, \beta, \delta) = \frac{1}{n} \left( \sum_{i=1}^{n} \frac{e^{\alpha + \frac{\delta}{\beta} X_i}}{1 + e^{\alpha + \frac{\delta}{\beta} X_i}} - \sum_{i=1}^{n} \frac{e^{\alpha + \beta X_i}}{1 + e^{\alpha + \beta X_i}} \right),$$

a linear approximation to $f(\alpha, \beta, \delta)$ expanded around the estimated value, $(\hat{\alpha}, \hat{\beta}, \hat{\delta})$, is given by
Figure A.1—The Prior Distribution for $\delta$ Implied by the Imputation of an Observation into the Likelihood Function

$$f_L(\alpha, \beta, \delta) = f(\hat{\alpha}, \hat{\beta}, \hat{\delta}) + \frac{\partial f}{\partial \alpha} (\alpha - \hat{\alpha}) + \frac{\partial f}{\partial \delta} (\delta - \hat{\delta})$$

$$+ \sum_{j=1}^P \frac{\partial f}{\partial \beta_j} (\beta_j - \hat{\beta}_j),$$

where

$$\frac{\partial f}{\partial \alpha} = \frac{1}{n} \left( \sum_{i=1}^n \left( \hat{P}_{H,i} (1 - \hat{P}_{H,i}) - \hat{P}_{N,i} (1 - \hat{P}_{N,i}) \right) \right),$$

$$\frac{\partial f}{\partial \delta} = \frac{1}{n} \sum_{i=1}^n \hat{P}_{H,i} (1 - \hat{P}_{H,i}),$$

$$\frac{\partial f}{\partial \beta_j} = \frac{1}{n} \left( \sum_{i=1}^n X_{i,j} \left( \hat{P}_{H,i} (1 - \hat{P}_{H,i}) - \hat{P}_{N,i} (1 - \hat{P}_{N,i}) \right) \right),$$

with the derivatives evaluated at \((\hat{\alpha}, \hat{\beta}, \hat{\delta})\).
The standard error is obtained by computing the variance of the linear approximation, \( f_L(\alpha, \beta, \delta) \), using the asymptotic variance matrix, \( \hat{\Sigma} \). Let zero reference the parameter \( \alpha \), and let \( X_{i,0} = 1 \). The contribution to the variance from \((\alpha, \beta)\) is

\[
\sum_{i=1}^{n} \sum_{j=1}^{n} \left( \hat{P}_{H,i}(1-\hat{P}_{H,i}) - \hat{P}_{N,i}(1-\hat{P}_{N,i}) \right) \left( \hat{P}_{H,j}(1-\hat{P}_{H,j}) - \hat{P}_{N,j}(1-\hat{P}_{N,j}) \right) X_i' \hat{\Sigma} X_j. \tag{A.5}
\]

The sum of the variance and covariance terms contributed by the \( \delta \) parameter is

\[
\left( \sum_{i=1}^{n} \hat{P}_{H,i} (1-\hat{P}_{H,i}) \right)^2 + 2 \sum_{i=1}^{n} \sum_{j=1}^{n} \hat{P}_{H,i} (1-\hat{P}_{H,i}) \hat{P}_{H,j} (1-\hat{P}_{H,j}) - \hat{P}_{N,j} (1-\hat{P}_{N,j}) X_j' \hat{\Sigma}_\delta, \tag{A.6}
\]

where \( \hat{\Sigma}_\delta \) is the vector of covariances of \( \delta \) with \((\alpha, \beta)\). The standard error, \( se_{mle} \), is obtained by adding Eqs. (A.5) and (A.6), dividing by \( n^2 \), and computing the square root of the resulting formula.

The same calculations can be reused to compute \( se_{b1} \). The only change required is that the predicted probabilities and asymptotic covariance matrices be replaced by the corresponding predicted values and covariance matrices based on the estimates \((\hat{\alpha}_{b1}, \hat{\beta}_{b1}, \hat{\delta}_{b1})\).
Appendix B

STANDARD ERRORS BASED ON REPEATED SAMPLING CONSIDERATIONS

Appendix B derives several alternative formulas for $se(\bar{Y} - \bar{P}_{N_{mle}})$, where $\bar{Y} - \bar{P}_{N_{mle}}$ estimates $\bar{F}_H - \bar{P}_N$ as discussed in Section 3. The estimate derived in this section is computed as part of the existing MMPS software.

The variance calculations performed here are based on a repeated sampling framework in which the unknown parameters are fixed, and the observed data are treated as realizations of random variables. Reviewing the survival models, the national death rate model is

$$\text{logit}(P_N(Y_i = 1 | X_i)) = \alpha + \beta X_i,$$

and the hospital-specific death rate model is

$$\text{logit}(P_H(Y_i = 1 | X_i)) = \alpha + \beta X_i + \delta.$$

The variance of the estimator is computed conditional on the values of the unknown parameters, $(\alpha, \beta, \delta)$. In addition, using the usual regression model assumptions, the covariates for the hospital, $X_1, \ldots, X_n$, and the covariates for the MMPS sample, $X_1^m, \ldots, X_n^m$, are held constant. Conditional on these parameters and covariates, the sample at the hospital whose death rate is being adjusted is collected independently of the MMPS sample, so that the 30-day mortality experience in the two samples is independent,

$$Y_1, \ldots, Y_n \perp Y_1^m, \ldots, Y_n^m.$$

This implies that

$$\bar{Y} \perp \bar{P}_{N_{mle}},$$

so that

$$\text{var}(\bar{Y} - \bar{P}_{N_{mle}}) = \text{var}(\bar{Y}) + \text{var}(\bar{P}_{N_{mle}}),$$

with no covariance term appearing in the variance of the difference in the two estimators. Daley et al. (1988, p. 3620) indicate that the covariance between the estimators is small and was omitted; in fact, it is exactly zero when computed within the repeated sampling framework.
An estimate of \( \text{var}(\hat{P}_{N_{\text{mle}}}) \) is obtained using standard methods. The same estimate of \( \text{var}(\hat{P}_{N_{\text{mle}}}) \) is used by both of the SE estimators proposed in this appendix. Let \( \hat{\Sigma}_{\text{mle}} \) represent the asymptotic covariance matrix of \( (\hat{\alpha}_{\text{mle}}, \hat{\beta}_{\text{mle}}) \) computed from the MMPS sample. Using the delta method to approximate \( \hat{P}_{N_{\text{mle}}} \) as a linear function of \( (\hat{\alpha}_{\text{mle}}, \hat{\beta}_{\text{mle}}) \) gives the following approximate variance for \( \hat{P}_{N_{\text{mle}}} \):

\[
\text{var}(\hat{P}_{N_{\text{mle}}}) = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\hat{e}_i \hat{e}_j}{(1 + \hat{e}_i)^2 (1 + \hat{e}_j)^2} X_i' \hat{\Sigma}_{\text{mle}} X_j,
\]

(B.1)

where \( \hat{e}_i \) is the estimated logit for the \( i \)th patient in the hospital sample computed using the national death rate model and \( (\hat{\alpha}_{\text{mle}}, \hat{\beta}_{\text{mle}}) \). This is the same formula as given by Daley et al. (1988) except that there is a typographical error in the exponential term in the numerator of their expression,

\[
e^{\hat{e}_i \hat{e}_j} \rightarrow e^{\hat{e}_i + \hat{e}_j}.
\]

Based on standard asymptotic theory, the sampling distribution of \( \hat{P}_{N_{\text{mle}}} \) is approximately normal.

The estimation of \( \text{var}(\bar{Y}) \) is complicated by the fact that this distribution depends on the unknown parameter, \( \delta \), and when simplifying model assumptions are not made, on the unknown hospital death probabilities, \( P_{H,i} \). One approach to estimating the \( \text{var}(\bar{Y}) \) is to set \( \delta = 0 \), the value corresponding to the null hypothesis that there are no differences in the quality of care between hospitals,

\[
\bar{P}_H = \bar{P}_N.
\]

Approximating \( \delta \) by zero,

\[
\text{var}(\bar{Y}) = \sum_{i=1}^{n} P_{H,i} (1 - P_{H,i})
\]

\[= \sum_{i=1}^{n} P_{N,i} (1 - P_{N,i}) .
\]

which can be estimated by

\[
\text{var}(\bar{Y}) = \sum_{i=1}^{n} \hat{P}_{N_{\text{mle}},i} (1 - \hat{P}_{N_{\text{mle}},i}) .
\]

(B.2)
This estimate should provide an adequate approximation except when \( \delta \) is large, because the variance of the binomial distribution, \( P_{H,i}(1-P_{H,i}) \), changes very slowly as a function of \( P_{H,i} \), and thus as a function of \( \delta \). The exception to this statement occurs when \( P_{H,i} < 0.10 \), because the binomial variance changes rapidly for death rates in this range. This variance estimate can also fail if the model for the national rates is very poor, but in this case, the potentially large biases in the estimate of \( \hat{P}_N \) are of much greater concern.

The distribution of \( \bar{Y} \) conditional on the covariates and the unknown parameters is also approximately normally distributed provided that a moderate (>20) number of patients in the hospital have death probabilities within the range (0.1,0.9), since the binomial variance is approximately constant for these rates so that a small number of patients do not dominate \( \bar{Y} \).

An estimate of the standard error of \( \bar{Y} - \hat{P}_{N,mle} \), which will be called the "conditional" standard error because it is computed by explicitly conditioning on the covariates in the hospital sample, is denoted by \( se_c \). It is obtained by combining the variance estimates in Eq. (B.1) and Eq. (B.2),

\[
(\text{se}_c)^2 = \text{var}(\bar{Y} - \hat{P}_{N,mle}) = \text{var}(\bar{Y}) + \text{var}(\hat{P}_{N,mle})
\]

\[
= \sum_{i=1}^{n} \hat{P}_{N,i}(1-\hat{P}_{N,i}) + \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{e_i^j + e_j^i}{(1+e_i^j)(1+e_j^i)^2} \hat{X}_i \hat{X}_j X_i X_j.
\]

This is the standard error currently employed by the MMPS.

**STANDARD ERROR BASED ON THE BINOMIAL APPROXIMATION**

An alternative estimate of \( \text{var}(\bar{Y}) \) that attempts to avoid the model dependence of the "conditional" standard error estimate of the variance is given by

\[
\text{var}(\bar{Y}) = \frac{\bar{Y}(1-\bar{Y})}{n}.
\]

This is the standard estimate of the unconditional (of \( X \)) variance of \( \bar{Y} \). Like the "conditional" variance estimate in the previous subsection, it is expected to provide an adequate approximation to the conditional variance of \( Y \) because of the fact that the binomial variance does not change much provided the death probabilities are not near the 0/1 boundaries. Unlike the estimate of \( \text{var}(\bar{Y}) \) contained in \( se_c \), Eq. (B.4) does not depend on the
national death rate model. Combining Eq. (B.4) with Eq. (B.1) gives the alternative “unconditional” standard error estimate, \( se_u \),

\[
(se_u)^{1/2} = \text{var} \left( \bar{Y} - \bar{P}_{N_mle} \right) \\
= \frac{\bar{Y}(1-\bar{Y})}{n} + \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{e^{i_j + i_j}}{(1+e^{i_j})^2(1+e^{i_j})^2} X_i^j \hat{\Sigma}_{mle} X_j.
\]

Note that this estimator also depends on the national death rate model through the estimate of \( \text{var} \left( \bar{P}_{N_mle} \right) \), but this dependence is slight because the \( \text{var}(\bar{Y}) \) is larger than \( \text{var} \left( \bar{P}_{N_mle} \right) \) by an order of magnitude for most hospitals.

The \( se_u \) will not reproduce the conditional SE exactly in large samples. Using a simple algebraic expansion, such as that given in HCFA (1989, p. A-19) yields

\[
\bar{P}_H(1-\bar{P}_H) = \frac{1}{n^2} \left( \sum_{i=1}^{n} P_{H,i}(1-P_{H,i}) + \sum_{i=1}^{n} (P_{H,i} - \bar{P}_H)^2 \right).
\]

Although this suggests that the \( se_u \) can be somewhat conservative for hospitals supplying large samples, a more serious difficulty with this estimator occurs with small samples that have no deaths. In this case, the estimate of \( \text{var}(\bar{Y}) \) is zero. Simulations in Appendix C show that the estimate is much too small in this situation.

**STANDARD ERROR CURRENTLY USED FOR HCFA REPORTING**

The standard error used to form the ranges in the HCFA mortality reports, denoted by \( se_{nu} \) (national unconditional), is similar to \( se_u \), except that instead of estimating \( \bar{P}_H \) using \( \bar{Y} \), \( \bar{P}_H \) is approximated by \( \bar{P}_{N_mle} \)

\[
se_{nu}^2 = \frac{\bar{P}_{N_mle}(1-\bar{P}_{N_mle})}{n} + \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{e^{i_j + i_j}}{(1+e^{i_j})^2(1+e^{i_j})^2} X_i^j \hat{\Sigma}_{mle} X_j.
\]  \( \text{(B.5)} \)

This standard error estimate has also been proposed by Dubois, Brook, and Rogers (1987). As shown by the discussion in the previous section, this SE is expected to be conservative except in those settings with \( P_{H,i} \ll P_{N,i} \).
The SE actually used by HCFA includes an additional positive term not given in Eq. (B.5). This term is intended to measure between-hospital variation in the death rates. It is derived from the national data for each HCFA medical condition. The data to reproduce this term for the MMPS medical conditions are not available, so this term will be omitted. Since it is always positive, it will tend to make $se_{HR}$ more conservative than reported in this Note.

The ranges reported by HCFA are based on additional skewness and kurtosis corrections. These corrections are also based on national data so they are also omitted in this Note.
Appendix C

SIMULATION STUDY OF STANDARD ERRORS FOR \( \bar{Y} - \bar{P}_{N\text{mle}} \)

Appendix C summarizes the results of the simulations described in Section 5 for the alternative standard errors, \( se_u \), \( se_{nu} \), \( se_{b1} \), and \( se_{pop} \). What follows gives more detail for each estimator and its SE. Table C.1 summarizes the coverage results by reporting various percentiles across the 180 settings in the simulation. The median coverage for the nominal 95 percent confidence intervals for the 180 simulation settings, along with the 5th and 95th percentiles of the simulated coverages, shows that the actual coverage probabilities are reasonably close to the nominal 95 percent level for each method for most of the simulation settings. At first blush, the Bayes-based estimator appears to offer a small improvement, but more careful review indicates some demanding settings where the Bayes procedure performs poorly.

COVERAGE PROBABILITIES

The simulation error for the estimated coverage probabilities is approximately 0.02. The final row of Table C.1 contains the estimated coverage probabilities using the correct population values to compute a SE. This gives an indication of the best coverage properties that can be expected.

Table C.2 provides the same summary statistics as Table C.1, except that the summary statistics are restricted to the 60 simulation settings with \( n = 25 \). The coverage probabilities are similar to the full simulation summary, except that \( se_u \) is now seen to perform very poorly in a substantial number of simulation settings. The Bayes-based procedure still appears to have slightly better coverage properties.

Table C.3 is the same as Table C.2 with the additional restriction that only simulation settings with the lowest severity are included in the summary statistics. These are the most demanding settings for the inferential procedures, since the estimates are near zero. There are no simulation settings near one.

The \( se_u \) produces intervals that are much too small for these simulation settings. The \( se_c \) based intervals have generally good performance but are noticeably too small for some of the simulation settings. The Bayes-based intervals also have even smaller coverages for many of these settings. The coverage probabilities for the intervals computed using the population-based SE perform surprisingly well considering the quality of the normal
Table C.1
Summary Statistics from All Simulation Settings

<table>
<thead>
<tr>
<th>Estimator</th>
<th>Median Coverage</th>
<th>Lowest 5% of Coverages</th>
<th>Highest 5% of Coverages</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\overline{Y} - \overline{P}_{\text{NMLE}}$</td>
<td>0.96</td>
<td>0.86</td>
<td>0.99</td>
<td>0.41</td>
<td>0.99</td>
</tr>
<tr>
<td>$se_u$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\overline{Y} - \overline{P}_{\text{NMLE}}$</td>
<td>0.95</td>
<td>0.88</td>
<td>0.99</td>
<td>0.86</td>
<td>0.99</td>
</tr>
<tr>
<td>$se_c$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\overline{Y} - \overline{P}_{\text{NMLE}}$</td>
<td>0.97</td>
<td>0.92</td>
<td>1.00</td>
<td>0.87</td>
<td>1.00</td>
</tr>
<tr>
<td>$se_{nu}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\overline{Y} - \overline{P}_{\text{Nb1}}$</td>
<td>0.95</td>
<td>0.91</td>
<td>0.98</td>
<td>0.80</td>
<td>0.99</td>
</tr>
<tr>
<td>$se_{b1}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\overline{Y} - \overline{P}_{\text{NMLE}}$</td>
<td>0.95</td>
<td>0.93</td>
<td>0.98</td>
<td>0.89</td>
<td>0.99</td>
</tr>
<tr>
<td>$se_{pop}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table C.2
Summary Statistics from Simulation Settings with $n = 25$

<table>
<thead>
<tr>
<th>Estimator</th>
<th>Median Coverage</th>
<th>Lowest 5% of Coverages</th>
<th>Highest 5% of Coverages</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\overline{Y} - \overline{P}_{\text{NMLE}}$</td>
<td>0.94</td>
<td>0.65</td>
<td>0.97</td>
<td>0.41</td>
<td>0.98</td>
</tr>
<tr>
<td>$se_u$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\overline{Y} - \overline{P}_{\text{NMLE}}$</td>
<td>0.95</td>
<td>0.88</td>
<td>0.98</td>
<td>0.86</td>
<td>0.99</td>
</tr>
<tr>
<td>$se_c$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\overline{Y} - \overline{P}_{\text{NMLE}}$</td>
<td>0.97</td>
<td>0.91</td>
<td>1.00</td>
<td>0.87</td>
<td>1.00</td>
</tr>
<tr>
<td>$se_{nu}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\overline{Y} - \overline{P}_{\text{Nb1}}$</td>
<td>0.95</td>
<td>0.90</td>
<td>0.98</td>
<td>0.80</td>
<td>0.98</td>
</tr>
<tr>
<td>$se_{b1}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\overline{Y} - \overline{P}_{\text{NMLE}}$</td>
<td>0.95</td>
<td>0.94</td>
<td>0.97</td>
<td>0.94</td>
<td>0.98</td>
</tr>
<tr>
<td>$se_{pop}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

approximation to the binomial distribution with the very small survival probabilities. For the most extreme of these settings, the good coverage properties do not result from the normal approximation, which is very poor, but may be an artifact of the simulation design that will fail if small variations on the current settings are studied.
Table C.3
Summary Statistics from Simulation Settings with \( n = 25 \) and Lowest Patient Severity

<table>
<thead>
<tr>
<th>Estimator</th>
<th>Median Coverage</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \bar{Y} - \bar{P}<em>{N</em>{mle}} ) ( se_u )</td>
<td>0.86</td>
<td>0.41</td>
<td>0.95</td>
</tr>
<tr>
<td>( \bar{Y} - \bar{P}<em>{N</em>{mle}} ) ( se_c )</td>
<td>0.96</td>
<td>0.86</td>
<td>0.98</td>
</tr>
<tr>
<td>( \bar{Y} - \bar{P}<em>{N</em>{mle}} ) ( se_{nu} )</td>
<td>0.97</td>
<td>0.87</td>
<td>1.00</td>
</tr>
<tr>
<td>( \bar{Y} - \bar{P}<em>{N</em>{b1}} ) ( se_{b1} )</td>
<td>0.92</td>
<td>0.80</td>
<td>0.96</td>
</tr>
<tr>
<td>( \bar{Y} - \bar{P}<em>{N</em>{mle}} ) ( se_{pop} )</td>
<td>0.95</td>
<td>0.94</td>
<td>0.98</td>
</tr>
</tbody>
</table>

The differences in the coverage probabilities of the different estimators suggest that the estimate of the \( \text{var}(\bar{Y}) \) is the dominant term in the estimate of the \( \text{var}(\bar{Y} - \bar{P}_{N_{mle}}) \), since each estimator shares a very similar estimate of \( \text{var}(\bar{P}_{N_{mle}}) \). The proportion of the \( \text{var}(\bar{Y} - \bar{P}_{N_{mle}}) \) due to \( \text{var}(\bar{P}_{N_{mle}}) \) is displayed in Figure C.1. Both variances are computed using population quantities. As anticipated, this proportion increases sharply as the hospital sample size increases. The dotplots also show that the proportion increases substantially as the severity of the case mix at the hospital increases. This fact was not obvious without simulation study since while it was known that \( \text{var}(\bar{P}_{N_{mle}}) \) is higher when the estimated death rates are higher and the severity variables are farther from their typical values, \( \text{var}(\bar{Y}) \) also increases as the severity increases. Most of the residual variation in the dotplots can be explained by variation in the simulated hospital quality, \( \delta \). \( \text{var}(\bar{Y} - \bar{P}_{N_{mle}}) \) increases with \( \delta \), while the numerator, \( \text{var}(\bar{P}_{N_{mle}}) \), is unchanged.

**POWER**

The coverage probabilities (and thus the size of the 5 percent test) were generally good in the null case with \( \delta = 0 \) for all of the SEs except \( se_u \). For all simulation settings with \( \delta \neq 0 \), the power of the 5 percent test was estimated from the simulation. Table C.4 shows the average power for each \( \delta \) and combination for simulation settings with moderate severity. The average power is computed over the four medical conditions. The simulation error for each power estimate obtained from Figure C.4 is approximately 0.06.
Figure C.1—Plots of \( \frac{\text{var}(\hat{P}_{\text{Nurl}})}{\text{var}(\bar{Y} - \hat{P}_{\text{Nurl}})} \) by Hospital Sample Size

Table C.4

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>SE Estimator</th>
<th>( \delta = -0.3 )</th>
<th>( \delta = 0.3 )</th>
<th>( \delta = -0.6 )</th>
<th>( \delta = 0.6 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>n = 25</td>
<td>( se_{b1} )</td>
<td>0.06</td>
<td>0.13</td>
<td>0.12</td>
<td>0.24</td>
</tr>
<tr>
<td>n = 50</td>
<td>( se_{b1} )</td>
<td>0.03</td>
<td>0.13</td>
<td>0.02</td>
<td>0.20</td>
</tr>
<tr>
<td>n = 100</td>
<td>( se_{nu} )</td>
<td>0.03</td>
<td>0.10</td>
<td>0.08</td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td>( se_{c} )</td>
<td>0.09</td>
<td>0.11</td>
<td>0.21</td>
<td>0.37</td>
</tr>
<tr>
<td></td>
<td>( se_{b1} )</td>
<td>0.04</td>
<td>0.12</td>
<td>0.22</td>
<td>0.29</td>
</tr>
<tr>
<td></td>
<td>( se_{nu} )</td>
<td>0.05</td>
<td>0.16</td>
<td>0.09</td>
<td>0.32</td>
</tr>
<tr>
<td>n = 100</td>
<td>( se_{c} )</td>
<td>0.16</td>
<td>0.20</td>
<td>0.44</td>
<td>0.60</td>
</tr>
<tr>
<td></td>
<td>( se_{b1} )</td>
<td>0.14</td>
<td>0.19</td>
<td>0.57</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td>( se_{nu} )</td>
<td>0.11</td>
<td>0.17</td>
<td>0.34</td>
<td>0.55</td>
</tr>
</tbody>
</table>
Very similar results were obtained for the high severity settings. The low severity simulation settings have markedly lower power (in the cells where the power is nontrivial), and the difference between the $\delta < 0$ and $\delta > 0$ settings is much more pronounced.

As anticipated, the principal factors influencing power are the sample size and the hospital effect size. The only other factor of consequence is the severity level; this factor contributes principally through the reduced power present at low severity levels (which is also reflected in the different medical conditions).

The $se_c$ estimator was almost uniformly better than the $se_{nu}$ estimator. This was expected because of the conservative nature of the $se_{nu}$ estimates. The $se_{b1}$ estimator had a much more mixed performance. It matched the performance of the $se_c$ estimator in many of the settings, but it was worse than both the $se_c$ and $se_{nu}$ estimator in other settings. No apparent trends in its differential performance were detected.

**RESULTS FOR THE $se_u$ ESTIMATOR**

The average of the unconditional variance estimates, $se_u^2$, was computed from the simulation replications at each simulation setting. The difference between the average estimated variance and the variance computed from the population values is displayed in Figure C.2. The $se_u$ overstates the variance for most settings as anticipated in Appendix B. Despite this fact, the intervals formed using $se_u$ tend to have less than the nominal 95 percent coverage, often much less than the stated precision. This appears to occur because the estimate needs some type of t-like correction because the variance is estimated.

The problem is most acute when $\bar{F}_H$ is near zero and the distribution of the estimated death rate is not approximately pivotal unless a very good variance estimate is available. The other problem that occurs when $\bar{F}_H$ is near zero is that $\bar{Y}$ is sometimes zero producing an estimate of the variance of $\bar{Y}$ equal to zero. The points in Figure C.2 with negative differences all correspond to settings with small $n$ and $\bar{F}_H$ near zero.

Figure C.3 displays the estimated coverages for $se_u$ plotted against the hospital death rates. The $se_u$ estimator was generally inferior; however, it performs very poorly only under the most demanding settings.

**RESULTS FOR THE $se_c$ ESTIMATOR**

The results for $se_c$ are given in Section 5.
Figure C.2—The Difference Between the Average (Across Simulation Replications) Variance Computed for $se_u$ and the Population-Based Variance vs the Population Variance

Figure C.3—Coverage Probabilities Estimated from the Simulation Using $se_u$ vs the Hospital Death Rate

NOTE: The highlighted points are the settings with $n = 25$ and the lowest severity
RESULTS FOR THE \(se_{nu}\) ESTIMATOR

From the discussion for \(se_c\), similar results are expected for \(se_{nu}\) except that \(se_{nu}\) tends to be larger in all cases. This is confirmed by Figure C.4, which shows that \(se_{nu}\) produces more than 95 percent coverage for most settings. The high-quality hospitals have variances that are overestimated because \(P_{N,i} > P_{H,i}\), and correspondingly for the low-quality hospitals. The conservative nature of the \(se_{nu}\) estimator improves its coverage probabilities for the settings where \(se_c\) tends to be too small, but it does so at the price of being too large for all other settings.

![Figure C.4](image)

Figure C.4—Coverage Probabilities Estimated for Each Simulation Setting Computed Using \(se_{nu}\) vs the National Death Rate for Patients Like Those Treated at the Hospital

RESULTS FOR THE \(se_{b1}\) ESTIMATORS

Table C.3 gives an overly pessimistic assessment of the performance of the \(se_{b1}\) estimator. Its performance is generally comparable to the SE based on the population-based variance except for two simulation settings. These settings are marked by x in Figure C.5.

These two points correspond to the simulation settings with \(\bar{P}_N = 0.038\), \(n = 25\), and \(\delta = -0.30\) (worse coverage = 0.80) and \(\delta = 0.30\) (coverage = 0.87). Since these points are intermediate to the simulation design space, the points with more extreme values of \(\delta\) would
be expected to be somewhat worse. The coverages for the same simulation settings but with $\delta = -0.60, 0.0, 0.60$ are 0.90, 0.93, and 0.90. These values suggest that the lowest coverage probability might be partly the result of simulation error and would be closer to 0.85 if the simulation were repeated.

It is interesting to note that the coverage probabilities for the estimates of $\delta$ are much better, between 0.95 and 0.98 for the five simulation settings discussed in the previous paragraph. The coverages for the estimates of $\delta$ are generally better than those for $\bar{P}_H - \bar{P}_N$. The median, 5 percent, 95 percent, minimum and maximum estimated coverages among all of the simulation settings for $\delta$ are 0.96, 0.94, 0.98, 0.89, and 1.0.

The fact that the coverage probabilities are better for $\delta$ than for $\bar{P}_H - \bar{P}_N$ suggests that improved numerical approximations to the posterior distribution (or likelihood function when the prior does not result in a proper posterior distribution) might substantially improve the intervals formed for $\bar{P}_H - \bar{P}_N$. 

---

Figure C.5—Coverage Probabilities Estimated for Each Simulation Setting Computed Using $se_{SI}$ vs the National Death Rate for Patients Like Those Treated at the Hospital
Appendix D
SAMPLE DESIGN FOR A FUTURE MMPS

Appendix D briefly summarizes the sample size and design discussion given in Longford (1991). The purpose of this appendix is to give an appropriate sampling design for estimating $\sigma_0^2$, and to suggest the possibility that a sample collected for other research purposes, such as estimating severity measure regression coefficients, could also be used to produce acceptably accurate estimates of $\sigma_0^2$ with only modest loss of precision for the regression coefficients of primary interest.

Longford (1991) obtained the following simple approximate formula for the information about $\sigma_0^2$ from a sample of $N$ patients with $n$ patients sampled per hospital:

$$\frac{1}{2} \frac{Nnw^2}{(1 + n\sigma_0^2)}$$

where $w = p(1-p)$ is the (assumed) common conditional variance given $\delta = 0$. For fixed values of $N$, $w$, and $\sigma_0^2$, the information is maximized when $n = 1/w\sigma_0^2$. For the medical conditions studied in this Note, $w = 0.15$, and it is likely that $\sigma_0^2 < 0.30$, suggesting a value of $n = 20$.

The experience of numerous researchers with clustered binary data is that $\sigma_0^2$ and $n$ must be very large before the standard errors of regression coefficients are substantially inflated by the amount of clustering in the PPS study design (Longford (1991) also has approximate formulas that affirm this finding). Thus, for only a modest increase in overall sample size (or corresponding sacrifice in precision), better estimates of $\sigma_0^2$ could be obtained from secondary analysis of an expensive-to-obtain sample in a study such as the PPS.
Appendix E
COMPARING THE INFORMATION FROM THE NATIONAL DISTRIBUTION WITH THE HOSPITAL-SPECIFIC INFORMATION

Appendix E provides calculations that allow the information obtained from national data sources summarized in the prior distribution to be compared to the information about a hospital obtained from the mortality rate and severity data at the hospital. The logit or δ scale is used for this comparison because a normal prior distribution (conditional on σ²) has been employed throughout, and for moderately large samples (200 is sufficient even for settings with the lowest death rates), the likelihood function is approximately normal as a function of δ. A natural comparison of the prior information about a hospital and the information obtained from the mortality rate at a hospital is between the variance of the prior, σ², and the variance of the asymptotic normal distribution approximating the likelihood function. Denote the latter variance by σ²^f.

The variance of the asymptotic normal approximation to the likelihood function is easy to compute, because for the logistic likelihood function given in Eq. (G.1), the asymptotic variance depends on the mortality data only through the estimates (\(\hat{\alpha}_{mle}, \hat{\beta}_{mle}, \hat{\delta}_{mle}\)). The asymptotic variance is computed with (\(\alpha, \beta\)) = (\(\hat{\alpha}_{mle}, \hat{\beta}_{mle}\)) along a grid of \(\hat{\delta}_{mle}\) values by inverting a weighted sum of squares matrix.

The σ²^f are evaluated on the hypothetical hospitals constructed for the simulation studies in Section 5. The low, medium, and high severity level mixes are reused to represent a very broad range of potential hospital patient mixes. The asymptotic variance of the likelihood function is also evaluated over the very broad range of potential hospital quality evaluated in Section 5, \(\hat{\delta}_{mle} = -0.60, -0.30, 0.00, 0.30, 0.60\). The σ²^f were computed for the largest hypothetical hospitals, which have 200 patients, in each of the medical conditions. Table E.1 displays the asymptotic variances from these hypothetical hospitals. Each variance reported in Table E.1 has been multiplied by 200, so for a similar hospital of size n, the asymptotic variance is obtained by the formula, σ²^f = (table entry)/n.

Two strong trends are apparent in Table E.1. For hospitals whose patients experience very low severity of condition, it is much more difficult to estimate hospital quality (i.e., σ²^f is much larger). Closely related to this trend is the fact that large positive estimates of δ are more accurate than large negative estimates of δ. An explanation of these trends and other related phenomena is given in Section 5. As discussed in Section 5, the low and high mixes of severity of patients’ conditions are at the outer limits of the plausible range.
Table E.1

The Asymptotic Variance, $\sigma^2$, of the Maximum Likelihood Estimate of $\delta$, with Each Entry Multiplied by 200 for a Similar Hospital of Size $n$, Table Entry = $N\sigma^2$

<table>
<thead>
<tr>
<th>Severity</th>
<th>$\hat{\delta}_{mle} = -0.6$</th>
<th>$\hat{\delta}_{mle} = -0.3$</th>
<th>$\hat{\delta}_{mle} = 0.0$</th>
<th>$\hat{\delta}_{mle} = 0.3$</th>
<th>$\hat{\delta}_{mle} = 0.6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stroke</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>24.9</td>
<td>20.4</td>
<td>16.9</td>
<td>14.1</td>
<td>12.1</td>
</tr>
<tr>
<td>Medium</td>
<td>12.6</td>
<td>10.8</td>
<td>9.4</td>
<td>8.4</td>
<td>7.6</td>
</tr>
<tr>
<td>High</td>
<td>9.1</td>
<td>8.4</td>
<td>7.9</td>
<td>7.5</td>
<td>7.2</td>
</tr>
<tr>
<td>Pneumonia</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>27.2</td>
<td>22.4</td>
<td>18.7</td>
<td>15.8</td>
<td>13.5</td>
</tr>
<tr>
<td>Medium</td>
<td>14.8</td>
<td>12.4</td>
<td>10.6</td>
<td>9.2</td>
<td>8.2</td>
</tr>
<tr>
<td>High</td>
<td>9.5</td>
<td>8.7</td>
<td>8.1</td>
<td>7.6</td>
<td>7.2</td>
</tr>
<tr>
<td>Myocardial Infarction</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>21.1</td>
<td>17.2</td>
<td>14.2</td>
<td>12.0</td>
<td>10.3</td>
</tr>
<tr>
<td>Medium</td>
<td>10.3</td>
<td>8.8</td>
<td>7.6</td>
<td>6.8</td>
<td>6.2</td>
</tr>
<tr>
<td>High</td>
<td>8.0</td>
<td>7.3</td>
<td>6.8</td>
<td>6.4</td>
<td>6.3</td>
</tr>
<tr>
<td>Congestive Heart Failure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>42.5</td>
<td>33.9</td>
<td>27.3</td>
<td>22.2</td>
<td>18.4</td>
</tr>
<tr>
<td>Medium</td>
<td>15.0</td>
<td>12.7</td>
<td>11.0</td>
<td>9.6</td>
<td>8.6</td>
</tr>
<tr>
<td>High</td>
<td>12.6</td>
<td>11.2</td>
<td>10.1</td>
<td>9.2</td>
<td>8.5</td>
</tr>
</tbody>
</table>

The information in the prior distribution for $\delta$ can be described in terms of the amount of equivalent data required so that $\sigma^2 = \sigma^2_0$. The equivalent hospital size for each of the simulated hospitals can be obtained from Table E.1 by the formula $n = \text{(table entry)} / \sigma^2_0$. Table E.2 displays the amount of data required from a hospital to match the prior information about a hospital in a typical setting with medium severity, for the range of maximum likelihood estimates given by $\hat{\delta}_{mle} = -0.30, 0.30$, for three different potential values of the mixing distribution, $\sigma_5 = 0.15, 0.25, 0.50$. The amount of equivalent data depends strongly on the variance of the mixing distribution. For most hospitals, the prior information will dominate the hospital-specific data unless the mixing distribution is much more diffuse than we anticipate; indeed, even if hospital sizes could be doubled or tripled (for example, by pooling mortality data across several years), the prior information would still be considerably stronger than the hospital-specific information for most hospitals.

Another way to summarize the relative information about a hospital from the prior distribution based on the national data and from the mortality data specific to the hospital is to compare the prior variance to the posterior variance. With the prior variance given by $\sigma^2_0$, and the likelihood approximated by a normal distribution with variance $\sigma^2_\delta$, Box and Tiao (1973) show that the posterior variance is given by $(1/\sigma^2_\delta + 1/\sigma^2_0)^{-1}$. Denoting the ratio of
Table E.2
The Sample Size at a Hospital Required to Match the Prior Information at the Hospital, with the Range Given for Each Entry Based on the Values $\hat{\delta}_{\text{mle}} = -0.30, 0.30$.

<table>
<thead>
<tr>
<th>Medical Condition</th>
<th>$\sigma_\delta = 0.15$</th>
<th>$\sigma_\delta = 0.25$</th>
<th>$\sigma_\delta = 0.50$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stroke</td>
<td>373-480</td>
<td>134-172</td>
<td>34-43</td>
</tr>
<tr>
<td>Pneumonia</td>
<td>409-551</td>
<td>147-198</td>
<td>37-50</td>
</tr>
<tr>
<td>Myocardial infarction</td>
<td>302-391</td>
<td>108-140</td>
<td>27-35</td>
</tr>
<tr>
<td>Congestive heart failure</td>
<td>427-564</td>
<td>154-203</td>
<td>38-51</td>
</tr>
</tbody>
</table>

$\sigma_\delta^2$ and $\sigma_\delta^2$ by $R = \sigma_\delta^2 / \sigma_\delta^2$, a simple derivation shows that the ratio of the posterior variance to the prior variance is given by $R/(R + 1)$. This implies, for example, that when the prior information and the information in the hospital's data are approximately equal, the data decrease our lack of uncertainty about the hospital by one-half. When the prior information is five times as strong as the information in the data (a very common situation under the more likely values of $\sigma_\delta^2$), the hospital-specific data reduce our uncertainty by only 17 percent.
Appendix F

COMPUTING MAXIMUM LIKELIHOOD ESTIMATES FOR THE LOGIT-NORMAL MODEL

This appendix describes the results of fitting models of the extrabinomial variability in hospital death rates. The logit-normal model fit to the hospital data is given in Eqs. (3.5) and (3.6) in Section 3. The log likelihood function for this model based on the PPS data is given by

\[
L(\alpha, \beta, \sigma_\delta^2) = \sum_{i=1}^{297} \log \int \left[ \prod_{j=1}^{n_i} \left( P_{H,ij}(\delta_i) \right)^{y_{ij}} \left( 1 - P_{H,ij}(\delta_i) \right)^{1-y_{ij}} \phi_{\sigma_\delta^2}(\delta_i) \right] d\delta_i , \tag{F.1}
\]

where \( P_{H,ij}(\delta_i) \) is

\[
P_{H,ij}(\delta_i) = \frac{e^{\alpha + \beta x_{ij}^P + \delta_i}}{1 + e^{\alpha + \beta x_{ij}^P + \delta_i}} ,
\]

and \( \phi_{\sigma_\delta^2} \) is the normal density with mean 0 and variance \( \sigma_\delta^2 \).

The same model, but without any covariates, was also fit to the national mortality data for each of the four medical conditions included in the MMPS for fiscal year 1986. This model, written in a slightly different form in Eqs. (3.3) and (3.4), was discussed in Section 3. Let

\[
Y_{i*}^T = n_i^T \overline{Y}_i^T
\]

where \( \overline{Y}_i^T \) is the Medicare population death rate at the \( i \)th hospital, and \( n_i^T, \ldots, n_T^T \) are the population sizes at each hospital in the nation for the condition. The log likelihood function specializes in this case to

\[
L(\alpha, \sigma_\delta^2) = \sum_{i=1}^{T} \log \int \left[ \left( P_H(\delta_i) \right)^{Y_{i*}^T} \left( 1 - P_H(\delta_i) \right)^{n_i^T - Y_{i*}^T} \phi_{\sigma_\delta^2}(\delta_i) \right] d\delta_i , \tag{F.2}
\]

where \( P_H(\delta_i) \) is the probability of death at the \( i \)th hospital,

\[
P_H(\delta_i) = \text{logit}^{-1}(\alpha + \delta_i) .
\]

The integrals in Eqs. (F.1) and (F.2) do not have closed-form representations; however, they can be approximated accurately and within computationally acceptable time limits using Gaussian quadrature. The results in this Note are based on nine-point quadrature,
which was found to be sufficient by comparing the results of nine-point quadrature to quadratures with differing numbers of points.

To obtain the MLE for the variance component, $\sigma^2_x$, and the other parameters, the likelihood functions were differentiated with respect to the parameters under the integrals (which are well behaved), yielding cumbersome, but computationally manageable formulas. The formulas for the first and second derivatives are given in Longford (1991), which also contains a computationally simple approximation to the likelihood function, and references to the methods used by other researchers to evaluate similar models. The integrals required to evaluate the derivatives have forms similar to those in Eqs. (F.1) and (F.2), and they were also evaluated using nine-point Gaussian quadrature.
Appendix G

EVALUATING THE MARGINAL POSTERIOR DISTRIBUTION OF $\delta$ AND $\bar{P}_H - \bar{P}_N$

In this appendix, the numerical methods used to evaluate the marginal posterior distribution of $\delta$ and $\bar{P}_H - \bar{P}_N$ are described, using the above prior with attention first focused on the evaluation of the marginal distribution of $\delta$. The prior distribution for $\delta$ has density

$$f_\nu(\delta) = \frac{\Gamma\left(\frac{1}{2}(\nu+1)\right)}{\left(\frac{\nu}{2}\right)^{\nu/2} \Gamma\left(\frac{\nu}{2}\right)} \frac{1}{\left(1+\frac{\delta^2}{\hat{\sigma}^2}\right)^{\nu/2}},$$

with $\nu = 3$. Combining $f_\nu$ with the log likelihood function (and the prior for $(\alpha, \beta)$) yields

$$l(\alpha, \beta, \delta) = -\frac{1}{2} \left( \begin{array}{c} \alpha - \hat{\alpha}_{mle} \\ \beta - \hat{\beta}_{mle} \end{array} \right)^T \hat{\Sigma}_{mle}^{-1} \left( \begin{array}{c} \alpha - \hat{\alpha}_{mle} \\ \beta - \hat{\beta}_{mle} \end{array} \right) + \sum_{i=1}^{n} (\alpha + \beta X_i + \delta) Y_i - \sum_{i=1}^{n} \log \left( 1 + e^{\alpha + \beta X_i + \delta} \right),$$

and integrating out $(\alpha, \beta)$ yields the marginal posterior distribution for $\delta$,

$$g(\delta) = \int f_\nu(\delta)e^{l(\alpha, \beta, \delta)} \, d\alpha \, d\beta.$$  \hspace{1cm} (G.1)

Because $(\alpha, \beta)$ is high dimensional, direct evaluation of the integral in Eq. (G.2) is difficult. The standard solution, based on asymptotic theory, is to approximate $f_\nu(\delta)e^{l(\alpha, \beta, \delta)}$ by a normal distribution with mean, $\left( \hat{\alpha}_b, \hat{\beta}_b, \hat{\delta}_b \right)$, at the mode of $f_\nu(\delta)e^{l(\alpha, \beta, \delta)}$, and covariance, $\hat{\Sigma}_b$, equal to negative the inverse of the Hessian of $f_\nu(\delta)e^{l(\alpha, \beta, \delta)}$. The values of $\left( \hat{\alpha}_b, \hat{\beta}_b, \hat{\delta}_b \right)$ maximizing $f_\nu(\delta)e^{l(\alpha, \beta, \delta)}$ are found using a Newton-Raphson routine. The normal approximation to the marginal distribution of $\delta$ is then obtained from the approximating multivariate normal distribution.

This asymptotic normal approximation is generally poor, especially when $l(\alpha, \beta, \delta)$ contains little information about $\delta$, or when $Y_i = 0$ for each subject, so that $l(\alpha, \beta, \delta)$ increases to an asymptote as $\delta \downarrow -\infty$. In the latter case, the behavior of the left tail of the posterior distribution is the same as that of the prior distribution, so the posterior mean and variance always exist, but the normal approximation to a $t_\nu$ distribution is very poor. In addition, the posterior distributions are typically skewed.
A much improved and computationally feasible method for evaluating $g(\delta)$ in Eq. (G.2) was developed in Tierney and Kadane (1986) which we use here. Briefly, the algorithm employed is: for a grid of $\delta$ values, $\delta_i$, $i = 1, \ldots, nqpt$,

1. find $\hat{a}_i, \hat{b}_i$ to maximize $l(\alpha, \beta, \delta_i)$,
2. approximate $g(\delta_i)$ by

$$
\hat{g}(\delta_i) = f_0(\delta_i)e^{l(\hat{a}_i, \hat{b}_i, \delta_i)} \frac{nqpt}{\sum_{j=1}^{nqpt} f_0(\delta_j)e^{l(\hat{a}_j, \hat{b}_j, \delta_j)}}.
$$

See Tierney and Kadane (1986) for a theoretical explanation of the advantages of this approach. Note that the determinant of the observed information for $\alpha, \beta$ conditional on $\delta$ is approximately constant and thus omitted here.

For our problem, this method appears to appropriately account for the skewness and kurtosis in the tails of the posterior distribution. We derive posterior means, percentiles, and quantiles from the grid of approximate density values using straightforward discrete density calculations and linear interpolation between grid points (when interpolation is required). Our simulation results support the adequacy of these numerical approximations.

Using the linear logistic survival model in Eq. (3.7), $\bar{F}_H - \bar{F}_N$ in Eq. (3.2) becomes

$$
\bar{F}_H - \bar{F}_N = \frac{1}{n} \sum_{i=1}^{n} \left\{ \logit^{-1}(\alpha + \beta X_i + \delta) - \logit^{-1}(\alpha + \beta X_i) \right\}.
$$

For fixed $(\alpha, \beta)$, $\bar{F}_H - \bar{F}_N$ is a monotone increasing function of $\delta$. For this derivation, let $X_{ij}, j = 1, \ldots, r$ be the $j^{th}$ severity covariate for the $i^{th}$ subject. The $(\alpha, \beta)$ are well determined from the large MMPS national sample. As a consequence, over the probable domain of $(\alpha, \beta)$, $\alpha + \beta X_i$ varies only modestly for each subject so that $\logit^{-1}$ is well approximated by a linear function applied to $\alpha + \beta X_i$ or $\alpha + \beta X_i + \delta$. Applying a linear approximation to $\logit^{-1}$ for each subject about some central value such as $\hat{a}_b, \hat{b}_b$ yields the following approximation to $\bar{F}_H - \bar{F}_N$ as a function of $(\alpha, \beta)$ for any fixed value of $\delta$,

$$
\bar{F}_H - \bar{F}_N = w_0(\alpha - \hat{\alpha}) + \frac{r}{j=1} w_j(\beta_j - \hat{\beta}_b,j),
$$

where

$$
w_j = \frac{1}{n} \sum_{i=1}^{n} X_{ij}(P_i(\delta)Q_i(\delta) - P_i(0)Q_i(0))
$$
with \( X_{i0} = 1 \) and

\[
P_i(\delta) = \logit^{-1}(\alpha + \beta X_i + \delta) \\
Q_i(\delta) = 1 - P_i(\delta)
\]

Unless \(|\delta|\) is large and many of the \( P_i(0) \) are very small, \(|w_j|\) are very small showing that \( \bar{P}_H - \bar{P}_N \) has a very weak dependence on \((\alpha, \beta)\). Hence, by choosing an appropriate value like the Bayes estimator, \( (\hat{\alpha}_b, \hat{\beta}_b) \) for \((\alpha, \beta)\), \( \bar{P}_H - \bar{P}_N \) can be approximated as a monotone function of \( \delta \), and the approximate marginal distribution of \( \bar{P}_H - \bar{P}_N \) can be obtained directly from the approximate discrete marginal distribution of \( \delta \).

By fixing \((\alpha, \beta) = (\hat{\alpha}_b, \hat{\beta}_b)\), we are not appreciably understating the uncertainty in \( \bar{P}_H - \bar{P}_N \) because we are appropriately representing the increased uncertainty in \( \delta \) because we do not know \((\alpha, \beta)\), and this uncertainty in \( \delta \) is in turn the primary source of uncertainty in \( \bar{P}_H - \bar{P}_N \). If the approximation introduced here were exact, the coverage properties of our confidence intervals for \( \delta \) and \( \bar{P}_H - \bar{P}_N \) would be identical. A comparison of the estimated coverage probabilities for \( \delta \) and \( \bar{P}_H - \bar{P}_N \) in the simulation settings described in Section 5 confirmed the adequacy of the approximation obtained by fixing \((\alpha, \beta) = (\hat{\alpha}_b, \hat{\beta}_b)\) in the calculation of \( \bar{P}_H - \bar{P}_N \). The average coverage probabilities for \( \bar{P}_H - \bar{P}_N \) were 0.005 and 0.02 less than the coverage probabilities for \( \delta \) in simulation settings with \( \delta = -0.60 \) and \( \delta = 0.60 \). The coverage probabilities were equal for less extreme \( \delta \) values.
Appendix H
SIMULATING NATIONAL MAXIMUM LIKELIHOOD ESTIMATES

As part of each replication, a simulated maximum likelihood estimate, \((\hat{\alpha}_{sim}, \hat{\beta}_{sim})\), based on the MMPS national sample was generated along with a simulated asymptotic covariance matrix, \(\hat{\Sigma}_{sim}\). Each replication used the same observed severity characteristics for each patient in the MMPS, \(X_1^m, ..., X_n^m\), but simulated the distribution of the mortality outcomes of the patients. Rather than actually generate the \(Y_i^m\) and compute the resulting estimates, however, the \((\hat{\alpha}_{sim}, \hat{\beta}_{sim})\) were generated directly from the approximate sampling distribution, \(N((\hat{\alpha}_{mle}, \hat{\beta}_{mle}), \hat{\Sigma}_{mle})\), obtained from the MMPS sample. The simulated asymptotic covariance matrix, \(\hat{\Sigma}_{sim}\), was then computed by evaluating the partial derivatives of the log likelihood function of the MMPS logistic regression at the value \((\hat{\alpha}_{sim}, \hat{\beta}_{sim})\), thus avoiding the generation of \(Y_i^m\) and computation of a logistic regression.

The simulation of \(((\hat{\alpha}_{mle}, \hat{\beta}_{mle}), \hat{\Sigma}_{mle})\) is intended to reflect the variability in these inputs into the estimators and is not intended as a check on the accuracy of the asymptotic approximations for the MMPS logistic regressions. Indeed, the simulation assumes that these approximations are exact. Because the MMPS sample sizes are large and the probabilities of death are not generally close to the 0/1 boundaries, the asymptotic approximations for the logistic regression should be very good.
Appendix I

CONSTRUCTING HYPOTHETICAL HOSPITALS WITH PATIENTS HAVING DIFFERENT MIXES OF SEVERITY

For each of the four medical conditions, hypothetical hospital samples were constructed that have different numbers of patients and different mixes of patient severity. The same approach was used for each medical condition.

Using the MMPS estimated mortality rate model, a predicted death probability was computed for each patient in the MMPS sample. Because the MMPS sampling design oversampled patients who died, selecting a random subsample from the MMPS was not appropriate even for producing the hospital with medium severity. To select patients with a specified severity mix, three equal-size strata (low, medium, high severity) of MMPS patients were formed based on the predicted death probabilities.

For hospitals with patients having a medium severity mix, patients were randomly selected from each stratum with the proportion of patients taken from each stratum chosen so that the expected value of the predicted probabilities from the stratified sample would be equal to the national death rate for the medical condition based on the national mortality data for fiscal year 1986. For example, for stroke patients, the stratum weights for the lowest to highest severity strata are (0.30, 0.56, 0.14), yielding an expected average death rate of 0.198. The same approach was used to construct the high-severity hospitals except that the proportion of patients chosen from each stratum was different. The stratum weights for stroke patients are (0.10, 0.37, 0.53), yielding an expected value for the predicted probabilities of 0.378. The severity mix of patients in the high-severity hospital is at the outer limits of the plausible range of hospital severity mix distribution based on the national mortality data as modelled in Section 3.

The patients for the low-severity hospitals were selected in a similar manner, but the algorithm had to be modified somewhat because the distribution of predicted death probabilities is very skewed toward high probabilities, with only a small proportion of patients having very low estimated death probabilities. To include more patients with low estimated death probabilities, the low-severity strata were divided into two substrata, and then four sampling proportions were selected for the four strata. For stroke patients, the sampling proportions are (0.60, 0.26, 0.10, 0.04) yielding an average estimated death probability of 0.092. It is apparent from the sampling proportions that a hospital treating patients like these would be extremely fortunate.
For the low, medium, and high conditions, stratified random samples of size 100 were selected. The hypothetical hospitals of size 25 and 50 were formed by taking the first 25 and 50 patients from the stratified sample of 100 patients. The patients for the smaller hospital sizes were not resampled for each replication of the simulation because the goal was to check that the estimation procedures are calibrated conditional on specific sets of severity characteristics, not just calibrated over the average of many different sets of severity characteristics.
Appendix J

DERIVATION OF THE EQUIVALENCE OF $\bar{P}_H - \bar{P}_N$ AND $\bar{Y} - \bar{P}_{N, mle}$

The log of the posterior distribution is given by

$$l(\alpha, \beta, \delta) = \log \left( f((\alpha, \beta, \delta)|Y_1, \ldots, Y_n, X_1, \ldots, X_n) \right) =$$

$$-\frac{1}{2} \left( \alpha - \hat{\alpha}_{mle} \right)^t \hat{\Sigma}_{mle}^{-1} \left( \alpha - \hat{\alpha}_{mle} \right) + \sum_{i=1}^{n} (\alpha + \beta X_i + \delta) Y_i - \sum_{i=1}^{n} \log \left( 1 + e^{\alpha + \beta X_i + \delta} \right).$$

(J.1)

The covariates measuring severity for the $i$th individual at a hospital are denoted by $X_{i1}, \ldots, X_{ir}$ in this derivation. The $(r + 1)$ estimating equations based on the partial derivatives of the log posterior with respect to the parameters $(\alpha, \beta)$ are given in matrix form by

$$0 = d - \hat{\Sigma}_{mle}^{-1} (\hat{\alpha} - \hat{\alpha}_{mle})$$

where

$$d = \begin{pmatrix}
\sum_{i=1}^{n} \hat{P}_{H,i} \\
\sum_{i=1}^{n} X_{i1} Y_i - \sum_{i=1}^{n} X_{i1} \hat{P}_{H,i} \\
\vdots \\
\sum_{i=1}^{n} X_{ir} Y_i - \sum_{i=1}^{n} X_{ir} \hat{P}_{H,i}
\end{pmatrix}.$$

Solving these equations yields

$$-\hat{\Sigma}_{mle}^{-1} (d) + \begin{pmatrix}
\hat{\alpha}_{mle} \\
\hat{\beta}_{mle}
\end{pmatrix} = \begin{pmatrix}
\hat{\alpha} \\
\hat{\beta}
\end{pmatrix}.$$

Thus, as $\|\hat{\Sigma}_{mle}\| \to 0$, $(\hat{\alpha}, \hat{\beta}) \to (\hat{\alpha}_{mle}, \hat{\beta}_{mle})$. This derivation is just a formal way of stating that the data from a single hospital do not update the precise prior information. The estimating equation based on the partial derivative of the log posterior with respect to $\delta$ is
\[ 0 = \sum_{i=1}^{n} Y_i - \sum_{i=1}^{n} \hat{P}_{H,i} , \]

which implies that

\[ \overline{\hat{P}}_H = \overline{Y} . \]

The fact that the sum of the estimated probabilities from a logistic regression is equal to the total number of cases with \( Y_i = 1 \) is well known (for example, Hosmer and Lemeshow (1989)). Thus, as the prior variance for \((\alpha, \beta)\) becomes increasingly sharp, \( \|\hat{P}_{mle}\| \to 0 , \)

\[ \overline{\hat{P}}_H - \overline{\hat{P}}_N \to \overline{Y} - \overline{P}_{N_{mle}} . \]
BIBLIOGRAPHY


Kahn, K. L., M. R. Chassin, L. V. Rubenstein, et al. (1988b), *Medical Record Abstraction Form and Guidelines for Assessing Quality of Care for Hospitalized Patients with Congestive Heart Failure*, RAND, N-2798-HCFA.


Keeler, E. B., K. L. Kahn, L. V. Rubenstein, et al. (1992), Hospital Characteristics Quality of Care and the HCFA Mortality Indicators, Submitted.


Roth, C. P., K. L. Kahn, M. J. Sherwood, et al. (1988), Medical Record Abstraction Form and Guidelines for Assessing Quality of Care for Hospitalized Patients with Pneumonia, RAND N-2801-HCFA.

Rubenstein, L. V., J. Kosecoff, K. L. Kahn, et al. (1988), Medical Record Abstraction Form and Guidelines for Assessing Quality of Care for Hospitalized Patients with Cerebrovascular Accident, RAND N-2802-HCFA.

Rubenstein, L. V., et al. (1991), Structured Implicit Review of the Medical Record: A Method for Measuring the Quality of Inpatient Medical Care and a Summary of Quality Changes Following Implementation of the Medicare Prospective Payment System, RAND N-3033-HCFA.

Sherwood, M. J., K. L. Kahn, J. Kosecoff, et al. (1988), Medical Record Abstraction Form and Guidelines for Assessing Quality of Care for Hospitalized Patients with Hip Fracture, RAND, N-2800-HCFA.
