Appendix A. Numeric example of Dimick Staiger Estimator and comparison between Dimick-Staiger Estimator and Hierarchical Poisson Estimator

As described in the manuscript, the Dimick-Staiger (DS) estimator has two components: volume-predicted mortality (hospitals' expected mortality rate based on their volume) and hospital-specific mortality. These two inputs are calculated for each hospital and are weighted based on the reliability of hospital-specific hospital mortality. For hospital *j*, the DS is calculated as:

$$\widehat{DS_j} = O_j \cdot W_j + \left(\widehat{\beta_0} + \hat{\beta}_1 \ln\left(Volume_j\right)\right) \cdot (1 - W_j)$$

Where *O* is hospital-specific mortality, β_0 and β_1 are estimated from a regression of mortality on hospital volume, and *W* is the weight assigned to hospital-specific mortality. In this equation, $\widehat{\beta_0} + \hat{\beta}_1 \ln\left(Volume_j\right)$ is volume predicted mortality.

The following is a numeric example showing the calculation of the DS Estimator in the context of the simulation.

A simulated admission-level dataset is created containing 3,000 hospitals and 315,601 admissions for one year of simulated admissions. First, hospital-specific mortality (O_j) is calculated for each hospital, which is simply the mean observed mortality rate given that patient risk is held constant across hospitals in the simulation. To obtain volume-predicted mortality, we estimate the following equation for hospital *j*:

$O_j = b_0 + b_1\, ln(volume)_j + e_j$

This is done using the following command in Stata:

```
reg mortmean ln_volume
```

```
      Source |       SS       df       MS              Number of obs =  315601
-------------+------------------------------           F(  1,315599) =28415.39
       Model |  274.001933     1  274.001933           Prob > F      =  0.0000
    Residual |  3043.23572315599  .009642729           R-squared     =  0.0826
-------------+------------------------------           Adj R-squared =  0.0826
       Total |  3317.23765315600  .010510892           Root MSE      =   .0982

------------------------------------------------------------------------------
   mortmean5 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   ln_sample |  -.0343134   .0002036  -168.57   0.000    -.0347124   -.0339144
       _cons |   .3312228    .001059   312.78   0.000     .3291473    .3332984
------------------------------------------------------------------------------
```

We then generate hospital-level predictions of mortality given each hospital's volume. Next, we require estimates of the weight (W_j) to apply to the hospital-specific and volume-predicted components of the DS. As shown in Appendix 1 of Dimick et al. 2009:

$$W_j = \frac{\hat{\sigma}^2}{\hat{\sigma}^2 + V_j}$$

Here, W_j is interpreted as the "ratio of signal variance to total variance…in the residual mortality rate" for hospital *j*.

To calculate $\hat{\sigma}^2$, we run the previously described regression model (`reg mortmean ln_volume`) and save the residuals ($\hat{e}_i$). We then calculate the variance of $\hat{e}$, weighting by the number of observations in each hospital, and call this scalar *tot_var*. In this example *tot_var* = .00879239.

Next, we need to calculate the amount of variance attributable to noise. To do this we estimate the following model for admission *i* at hospital *j*:

$Mortality_{ij} = b_j u_j + e_{ij}$

Where u_j is a vector of hospital fixed effects. This is implemented in Stata using:

```
Linear regression, absorbing indicators                Number of obs =  315601
                                                       F( 0, 312601) =       .
                                                       Prob > F      =
                                                       R-squared     =  0.0802
                                                       Adj R-squared =  0.0714
                                                       Root MSE      =   .3489

------------------------------------------------------------------------------
       death |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   .1551643   .0006211   249.84   0.000      .153947    .1563815
-------------+----------------------------------------------------------------
          id |   F(2999, 312601) =      9.086   0.000        (3000 categories)

ereturn list

e(rmse) =  .34890
```

We calculate the mean-squared error (*MSE*) from the equation: MSE = .12173463. We then calculate a hospital-specific estimate of noise (V_j in the above equation) by dividing the MSE value by the number of observations in the hospital:

$$V_j = \frac{MSE}{Obs_j}$$

Finally we calculate the mean of V_j for all hospitals ($\bar{V}$=.0005278), weighting by the number of observations in each hospital. Then,

$\hat{\sigma}^2$= *tot_var* - $\bar{V}$ =.00879239- .0005278 = .00826463

We have now defined, $\hat{\sigma}^2$ and V_j, allowing us to calculate W_j and implement the DS estimator.

In this example there is a hospital that has 10 observations, where V_j=.0121735, RA mortality=.30, and volume predicted mortality=.2522134. For this hospital, W_j = .404373, and the DS estimator is calculated as:

$$\widehat{DS_j} = .30 \cdot .404373 + .2522134 \cdot (1 - .404373) = .27153701$$

In this example, there is another hospital that has 893 observations, where V_j= .0001363, RA mortality = .0515118, and volume predicted mortality = .0980776. For this hospital, W_j = .9837732, and the DS estimator is calculated as:

$$\widehat{DS_j} = .0515118 \cdot .9837732 + .0980776 \cdot (1 - .9837732) = .0522674$$

B.1: Hierarchical Poisson (HP) Model Assumptions

Consider the following hierarchical model: for m independent units (e.g., hospitals), let

- $Y_i|\theta_i \sim \text{Poisson}(e_i\theta_i)$, where $e_i > 0$ is a *known* quantity.
- $\theta_i \sim \text{Gamma}(\delta, \beta_i)$, parameterized such that $E(\theta_i) = \frac{\delta}{\beta_i}$, $Var(\theta_i) = \frac{\delta}{\beta_i^2}$, where $\delta > 0$ and $\beta_i > 0,\ i = 1 \ldots m$.
- $\mu_i = \frac{\delta}{\beta_i}$; that is, $E(\theta_i) = \mu_i$. Observe that $\beta_i = \frac{\delta}{\mu_i}$.

Under this hierarchical model, easy computations (e.g., see Lawless, 1987 Canadian Journal of Statistics) immediately establish the following two facts:

1. The marginal probability distribution of Y_i (i.e., integrating out θ_i) is given by

$$P(Y_i = y_i) = \frac{\Gamma(y_i + a^{-1})}{y!\Gamma(a^{-1})} \left(\frac{a\mu_i}{1 + ae_i\mu_i}\right)^{y_i} \left(\frac{1}{1 + ae_i\mu_i}\right)^{a^{-1}} \tag{1}$$

   where $a = \delta^{-1}$. This fact will be important later and, in particular, allows us to make direct use of Stata for computation.

2. The conditional density of θ_i, given $Y_i = y_i$, is Gamma$(y_i + \delta, e_i + \beta_i)$. Hence,

$$E(\theta_i|Y_i = y_i) = \frac{y_i + \delta}{e_i + \beta_i}.$$

   Recalling that $\beta_i = \frac{\delta}{\mu_i}$, this last result can also be written as follows:

$$E(\theta_i|Y_i = y_i) = \frac{y_i}{e_i}(1 - B_i) + B_i\mu_i, \tag{2}$$

   where

$$B_i \ = \ \frac{\frac{\delta}{\mu_i}}{e_i + \frac{\delta}{\mu_i}} \ = \ \frac{\delta}{\mu_i e_i + \delta}.$$

   Letting $\hat{\theta}_i = \frac{y_i}{e_i}$, we have

$$E(\theta_i|Y_i = y_i) = \hat{\theta}_i(1 - B_i) + B_i\mu_i, \tag{3}$$

   providing the basis for computing a Stein-like shrinkage estimator for θ_i: a weighted linear combination of the empirical estimate $\hat{\theta}_i$ and $\mu_i = E(\theta_i)$.

B.2: Estimation via Negative Binomial Regression

Suppose that $\mu_i = \mu_i(b)$, where $\log \mu_i(b) = x_i'b$, where x_i includes a '1' for the intercept; that is, consider a "log" link function. Using this parameterization ensures that $\mu_i > 0$ and, usefully, we can estimate the regression parameter b without any constraints.

As indicated earlier in (1), the marginal probability distribution of Y_i under this HP model is given by

$$P(Y_i = y_i) = \frac{\Gamma(y_i + a^{-1})}{y!\Gamma(a^{-1})} \left(\frac{a\mu_i}{1 + ae_i\mu_i}\right)^{y_i} \left(\frac{1}{1 + ae_i\mu_i}\right)^{a^{-1}}$$

where $a = \delta^{-1}$. Taking the natural logarithm and using the fact that $\log \mu_i(b) = x_i'b$, one obtains the following loglikelihood for a and b (see, eg, Lawless 1987, bottom of page 210)

$$\ell(b, a) = c(y_1 \dots y_n) + \sum_{i=1}^{m} \left[\sum_{j=0}^{y_i - 1} \log(1 + aj)\right] + y_i \log(e_i\mu_i(b)) - (y_i + a^{-1}) \log(1 + ae_i\mu_i(b)),$$

where $c(y_1 \dots y_n)$ is a constant term that doesn't depend on a or b and the summation term in $[\cdot]$ is zero if $y_i \leq 1$.

One can show that this loglikelihood is *exactly* the same as that for a negative binomial regression model: $Y_i \sim \text{NegBin}(e_i\mu_i(b), a)$, $i = 1 \dots n$, in which $E(Y_i) = e_i\mu_i(b)$ and $Var(Y_i) = e_i\mu_i(b)(1 + ae_i\mu_i(b))$. If $a = 0$ (no overdispersion), this reduces to a Poisson regression model. It is now possible to estimate a and b as described in Lawless (1987).

Under the log-link parameterization $\log \mu_i(b) = x_i'b$, the mean count can also be rewritten as $E(Y_i) = \exp(v_i + x_i'b)$, where $v_i = \log e_i$ is an *offset*. If we take $e_i = 1, i = 1 \dots m$, there is no offset ($v_i = 0,\ i = 1 \dots m$). This regression model is the default model fit in Stata using the `nbreg` command. (Note: if there is an offset, the `offset` option can be used to include one; see the discussion at the end of Appendix B).

Consider a hospital-level dataset with 3000 hospitals, each with a designated volume of patients and number of deaths occurring within 30 days of admission (deathcount). We first re-parameterize volume, taking its natural logarithm; call this w (i.e., $w = \log vol$). To implement the HP model in Stata (i.e., without an offset, or $e_i = 1$ for each i), we use the following command:

```
nbreg deathcount w
```

Output:

```
nbreg deathcount w

Negative binomial regression                          Number of obs   =       3000
                                                      LR chi2(1)      =    8133.01
Dispersion     = mean                                 Prob > chi2     =     0.0000
Log likelihood = -8183.0165                           Pseudo R2       =     0.3320

------------------------------------------------------------------------------
 deathcount  |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           w |   1.009213    .006639   152.01   0.000     .9962004    1.022225
       _cons |  -1.650409   .0324994   -50.78   0.000    -1.714107   -1.586711
-------------+----------------------------------------------------------------
    /lnalpha |  -3.481276   .0597869                     -3.598456   -3.364096
-------------+----------------------------------------------------------------
       alpha |   .0307681   .0018395                      .0273659    .0345933
------------------------------------------------------------------------------
Likelihood-ratio test of alpha=0:  chibar2(01) =  905.28 Prob>=chibar2 = 0.000
```

Here, $b = (b_0, b_1)'$; under the `Coef.` column, `_cons` is the estimated value of b_0 and `w` is the estimated value of b_1. Importantly, `alpha` equals the estimated value of a (i.e. δ^{-1}) in the Negative Binomial model parameterization given earlier.

The output above further implies that our estimate of μ_i is given by

$$\hat{\mu}_i = \mu_i(\hat{b}) = \exp(-1.650409 - 1.009213\ w_i).$$

Using (3), the corresponding Hierarchical Poisson estimator for the mean 30-day mortality count in hospital i (i.e., with $e_i = 1$) is given by

$$\hat{\theta}_i = y_i(1 - \hat{B}_i) + \hat{B}_i\ \hat{\mu}_i,$$

where

$$\hat{B}_i\ =\ \frac{\hat{\delta}}{\hat{\mu}_i + \hat{\delta}}$$

Using the output above, $\hat{\delta} = 1/\texttt{alpha} = 1/.0307681 = 32.50$; therefore,

$$\hat{B}_i\ =\ \ =\ \frac{32.50}{\hat{\mu}_i + 32.50}.$$

Rescaling $\hat{\theta}_i$ by hospital volume (vol_i), one obtains

$$\hat{\varphi}_i = \frac{\hat{\theta}_i}{vol_i},$$

an estimate of the hospital-specific 30-day mortality rate. This can also be estimated directly with `nbreg` as above, using the offset $e_i = vol_i$.

B.3 Extensions to Modeling Rate-Adjusted Mortality

In this paper, Y_i represents the death count in hospital i in a fixed time period and μ_i depends on a fixed set of covariates x_i (e.g., hospital volume in this same period). In the absence of an offset (i.e., $e_i = 1$, $i = 1 \ldots m$), the conditional Poisson mean θ_i thus represents the hospital-specific mean death count for the i^{th} hospital; similarly, μ_i is the expected death count for a hospital with characteristics x_i. For this reason,

More generally, and as can be seen directly in (2) and (3), it is possible use alternative forms of standardization to change interpretation of θ_i from a count to a standardized rate. In particular, we may set e_i to be equal to a "known" measure of the expected number of deaths in hospital i (e.g., obtained using some form of internal or external standardization). Then, θ_i can be interpreted as a hospital-specific standardized mortality ratio (i.e., dependent on a hospital-specific random effect and dependent on the standardization used to calculate e_i); similarly, μ_i is the expected SMR for a hospital with characteristics x_i.

This second approach is comparable to what is done for the unadjusted DS estimator, where the adjustment for hospital volume (x_i) can be viewed as an indirect form of standardization and one takes $e_i = n_i$, the number of patients served by hospital i. We refer to this as the unadjusted HP model estimator. In this case, the estimated value of θ_i is shrinkage estimator of the probability of an event (e.g., 30 day mortality). More generally, a measure of rate-adjusted mortality (or adjusted HP model-based estimator) can easily be obtained analogously to the DS estimator upon estimating the regression relationship between the average hospital death count and hospital volume and then substituting this estimated value in for e_i.

It is important to note in closing that, as presently described, the HP model only makes use of information aggregated at the hospital level and, in particular, uses the parametric relationship between the mean and variance that is induced under this model in order to estimate the variance components necessary for computing (3).

Comparing the unadjusted HP and DS estimators

As will be shown below, the main difference between the unadjusted HP model-based estimator and the unadjusted DS estimator stems from the methodology used to estimate the shrinkage factor B_i (i.e., assuming the same parameterization is used for μ_i). It is noted here that the DS estimator, as originally proposed, also parameterizes the mean of the "structural quality" distribution as $\mu_i = \beta_0 + \beta_1 \log(volume)$; in contrast, the HP estimator assumes $\log \mu_i = \beta_0 + \beta_1 \log(volume)$. However, other parameterizations can be used for either estimator; using the same parameterization for μ_i in both leads one back to comparing shrinkage factors alone.

The HP model is arguably most appropriate when the number of events in each hospital is small in comparison to the overall volume of patients; otherwise, a mixed effects logistic or binomial model might be a better choice. We have

$$B_i \;=\; \frac{\frac{\delta}{\mu_i}}{e_i + \frac{\delta}{\mu_i}} \;=\; \frac{\delta}{\mu_i e_i + \delta}.$$

Estimation of B_i is possible using maximum likelihood (ie, via the negative binomial marginal distribution for Y_i). We now show that B_i depends on two components of variance. Since $\mu_i = \delta_i/\beta_i$, algebra shows

$$B_i \;=\; \frac{\mu_i e_i}{\mu_i e_i + e_i^2 \frac{\mu_i}{\beta_i}}.$$

Since $Y_i|\theta_i \sim \text{Poisson}(e_i\theta_i)$, it follows that $Var(Y_i|\theta_i) = e_i\theta_i$; since $E(\theta_i) = \frac{\delta}{\beta_i} = \mu_i$, we may now write $\mu_i e_i = E(Var(Y_i|\theta_i))$. Similarly, we may write

$$e_i^2 \frac{\mu_i}{\beta_i} = e_i^2 \frac{\delta}{\beta_i^2} = e_i^2 Var(\theta_i).$$

Since $e_i^2 Var(\theta_i) = Var(e_i\theta_i) = Var(E(Y_i|\theta_i))$, we have now shown

$$B_i \;=\; \frac{E(Var(Y_i|\theta_i))}{E(Var(Y_i|\theta_i)) + Var(E(Y_i|\theta_i))} \;=\; \frac{E(Var(Y_i|\theta_i))}{Var(Y_i)}.$$

The unadjusted DS estimator is similar to that in Morris (1983 JASA), who uses a hierarhically specified normal model to motivate the form of the shrinkage estimator. Appendix 1 of Dimick et al. (2009) shows that the original estimator of Morris (Section 5) is not used; instead a non-iterative procedure is employed, where V_i has been replaced by an estimator that assumes that all subjects within a hospital have the same (conditional) mean and variance. The latter estimator requires the existence of replications and, in addition, asserts that no subject-level heterogeneity in mean or variance exists within a hospital.

We now demonstrate that the main difference between the Hierarchical Poisson and DS estimators relates to the method by which the variance components are estimated. The

unadjusted DS model assumes $Y_i|\theta_i \sim \text{Normal}(\theta_i, V_i)$ and $\theta_i \sim \text{Normal}(\mu_i, A_i)$. In this case, the relevant shrinkage factor is given by

$$B_i^{(DS)} = \frac{V_i}{A_i + V_i} = \frac{Var(Y_i|\theta_i)}{Var(Y_i|\theta_i) + Var(\theta_i)}.$$

However, in this normal model, $Var(Y_i|\theta_i)$ doesn't depend on θ_i. As a result, $Var(Y_i|\theta_i) = E(Var(Y_i|\theta_i))$ and it follows that

$$B_i^{(DS)} = \frac{E(Var(Y_i|\theta_i))}{E(Var(Y_i|\theta_i)) + Var(E(Y_i|\theta_i))} = \frac{E(Var(Y_i|\theta_i))}{Var(Y_i)}.$$

Written in this way, $B_i^{(DS)}$ is observed to be identical to B_i above, in that it involves exactly the same two variance components.

Consequently, the HP and DS estimators differ primarily in the assumptions and methods used to estimate the two variance components appearing in the shrinkage factor. In particular, and as described earlier, the HP model only makes use of information aggregated at the hospital level and uses the parametric relationship between the mean and variance that is induced under this model in order to estimate these two variance components. In contrast, the DS estimator uses a moment-based approach for estimation and does not assume the existence a parametric relationship between mean and variance. Instead, as described above, the DS estimator introduces the alternative assumption that all subjects within a hospital share a common conditional mean and an unrelated common conditional variance, and utilizes individual-level information in order estimate both variance components.

Appendix C. Volume and Mortality Relationship

Figure C1. Volume and Mortality Relationship Estimated from CMS Data and Implemented in Simulation

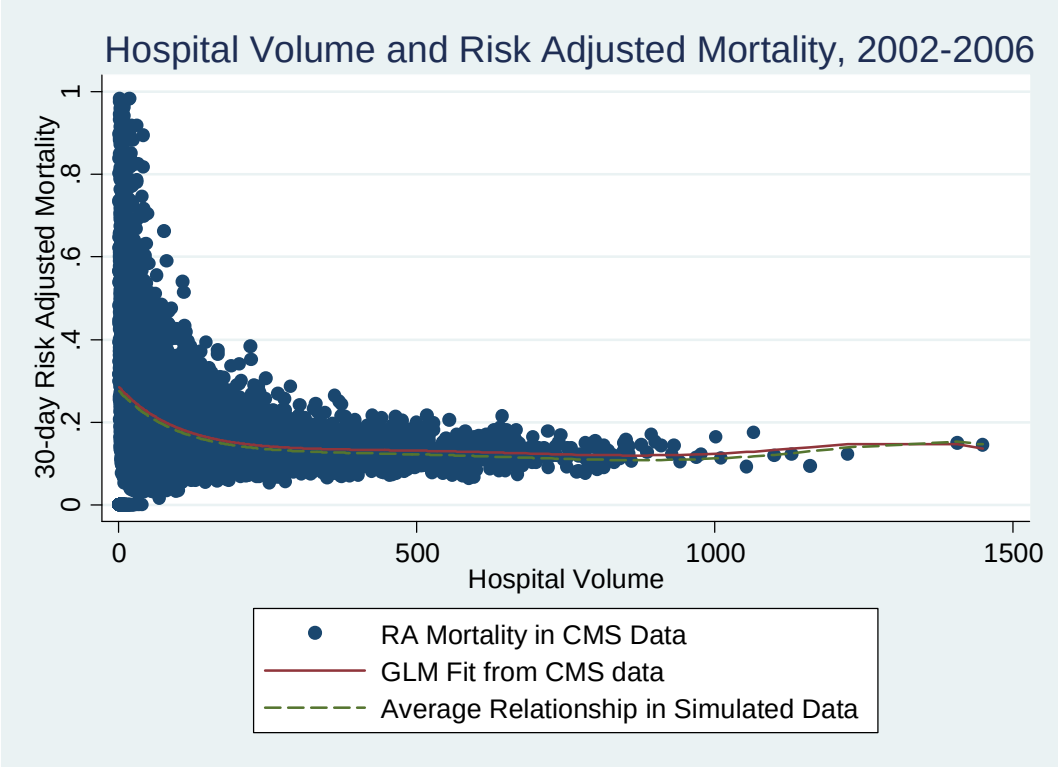


Figure C1, shows the volume and 30-day mortality relationship using Medicare inpatient data from 2002-2006 for 16,065 hospital-year observations and the average relationship in the simulated data. The red line shows the predicted relationship between volume and 30-day mortality, estimated from a generalized linear model with a binomial family and logit link function. We specified a 5th degree polynomial function (*volume, volume*^{2}*, volume*^{3}*, volume*^{4}*, volume*^{5}), which fits the observed data relatively well. This estimated relationship was then specified in the simulation model, as shown in the dashed green line, which is very similar to the relationship estimated with the CMS data. The relationship in the simulated data was derived by calculating the average coefficient for each volume term in the generalized linear model (estimated for each simulation iteration), averaging the coefficients across the 1,000 iterations, and then plotting this function over hospital volume.

Appendix D. Data generating functions for simulation

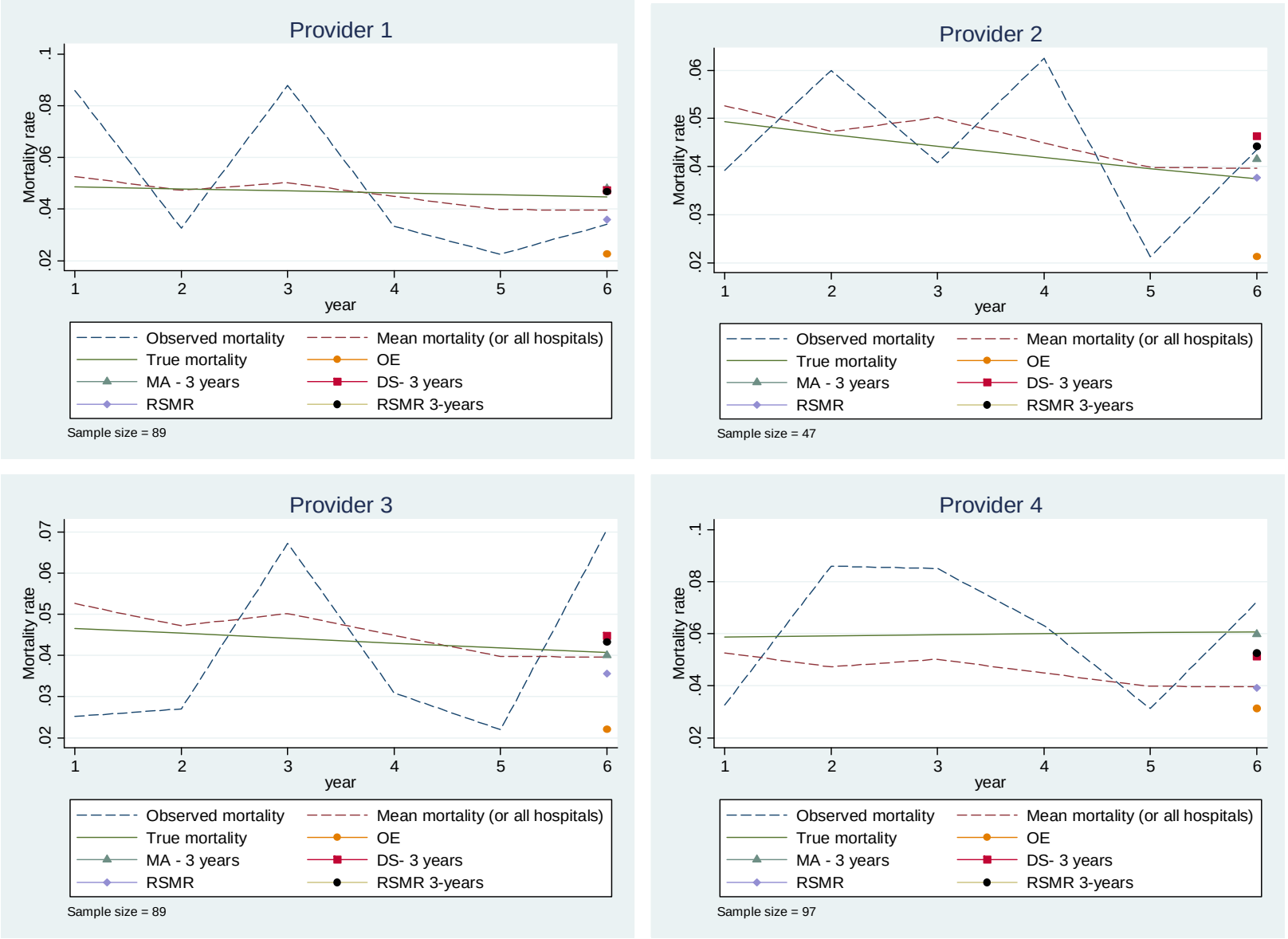
Table D1. Data generating functions for simulation

Data Element	Data Generation Process	Parameter values
Sample size in first year	$Sample\ size_{i1} \sim Gamma(k, \theta) * Z \quad if \quad 1450 >= X >= 1$ 1 if X =0 1450 if X > 1450	k=1.01 θ=4.65 Z=22.12
Sample size change	$Sample_{it} = Sample_{it-1} * C$ $C \sim N(\mu, \sigma)$	μ = 1.001 σ = .06
Raw true mortality score in year 1	$Mortality\ raw_{i1} \sim N(\mu, \sigma) \quad if \quad 1 >= X >= 0.005$ 0.005 if X < 0.005 1 if X > 1	μ =.23813 σ =.032
Volume adjusted true mortality score in year 1†	$Mortality\ adjusted_{i1} = Mortality\ raw_{i1} + volume\ adjustment_{i1} \quad if \quad 1 >= X >= 0.005$ 0.005 if X < 0.005 1 if X > 1	$volume\ adjustment_{it} = \left(\frac{\exp(bX_{it})}{1+\exp(bX_{it})}\right) - \left(\overline{\frac{\exp(bX_{it})}{1+\exp(bX_{it})}}\right)$ $bX_{it} = -0.913 - (7.76\text{ E-}3 * volume_{it}) + (2.49\text{ E-}5 * volume^2_{it}) - (3.86\text{ E-}8 * volume^3_{it}) + (2.78\text{ E-}11 * volume^4_{it}) - (7.36\text{ E-}15 * volume^5_{it})$
Volume adjusted true mortality score in years 2-4	$Mortality\ adjusted_{it} = (Mortality\ raw_{it-1} + volume\ adjustment_{it}) * K_i^{\ t} \quad if \quad 1 >= X >= 0.005$ $K_i \sim N(\mu, \sigma)$ 0.005 if X < 0.005 1 if X > 1	μ =.956 σ =.057

Note: *i* indexes to hospital and *t* indexes to year of simulation.
†The coefficients in *b* vector for *volume adjustment* were estimated using a generalized linear model from the binomial family with a logit link. Note that, as a result of subtracting the mean *volume adjustment*, $volume\ adjustment_{it}$ has a mean of 0.

Appendix E. Visual depiction of estimators

Figure E1. Hypothetical estimates of mortality quality for four providers using alternative estimators



In Figure E1, the X-axis is time and the Y-axis is hospital mortality. The solid green line is true mortality for a given hospital, the dashed blue line is observed mortality for that hospital, and the dashed red line is the mean observed mortality for all hospitals. Figure E1 shows that the alternative estimators can arrive at substantially different estimates of true mortality performance, estimates that can vary substantially from true performance.