AHRQ Quality Indicators

Agency for Healthcare Research and Quality
AHRQ Quality Indicators (AHRQ QI)

Risk Adjustment and Hierarchical Modeling Workgroup
Final Report
January 19, 2007

AHRQ Quality Indicators
Risk Adjustment and Hierarchical Modeling Approaches

1 Introduction

The Inpatient Quality Indicators (IQIs) are a set of measures that provide a perspective on hospital quality of care using hospital administrative data. These indicators reflect quality of care inside hospitals and include inpatient mortality for certain procedures and medical conditions; utilization of procedures for which there are questions of overuse, underuse, and misuse; and volume of procedures for which there is some evidence that a higher volume of procedures is associated with lower mortality.

The IQIs are a software tool distributed free by the Agency for Healthcare Research and Quality (AHRQ). The software can be used to help hospitals identify potential problem areas that might need further study and which can provide an indirect measure of inhospital quality of care. The IQI software programs can be applied to any hospital inpatient administrative data. These data are readily available and relatively inexpensive to use.

Inpatient Quality Indicators:

- Can be used to help hospitals identify potential problem areas that might need further study.
- Provide the opportunity to assess quality of care inside the hospital using administrative data found in the typical discharge record.
- Include 15 mortality indicators for conditions or procedures for which mortality can vary from hospital to hospital.
- Include 11 utilization indicators for procedures for which utilization varies across hospitals or geographic areas.
- Include 6 volume indicators for procedures for which outcomes may be related to the volume of those procedures performed.
- Are publicly available without cost , and are available for download

The IQIs include the following 32 measures:

1. Mortality Rates for Medical Conditions (7 Indicators)
    - Acute myocardial infarction (AMI) (IQI 15)
    - AMI, Without Transfer Cases (IQI 32)
    - Congestive heart failure (IQI 16)
    - Stroke (IQI 17)
    - Gastrointestinal hemorrhage (IQI 18)
    - Hip fracture (IQI 19)
    - Pneumonia (IQI 20)
2. Mortality Rates for Surgical Procedures (8 Indicators)

- Esophageal resection (IQI 8)
- Pancreatic resection (IQI 9)
- Abdominal aortic aneurysm repair (IQI 11)
- Coronary artery bypass graft (IQI 12)
- Percutaneous transluminal coronary angioplasty (IQI 30)
- Carotid endarterectomy (IQI 31)
- Craniotomy (IQI 13)
- Hip replacement (IQI 14)

3. Hospital-level Procedure Utilization Rates (7 Indicators)
    - Cesarean section delivery (IQI 21)
    - Primary Cesarean delivery (IQI 33)
    - Vaginal Birth After Cesarean (VBAC), Uncomplicated (IQI 22)
    - VBAC, All (IQI 34)
    - Laparoscopic cholecystectomy (IQI 23)
    - Incidental appendectomy in the elderly (IQI 24)
    - Bi-lateral cardiac catheterization (IQI 25)
4. Area-level Utilization Rates (4 Indicators)
    - Coronary artery bypass graft (IQI 26)
    - Percutaneous transluminal coronary angioplasty (IQI 27)
    - Hysterectomy (IQI 28)
    - Laminectomy or spinal fusion (IQI 29)
5. Volume of Procedures (6 Indicators)
    - Esophageal resection (IQI 1)
    - Pancreatic resection (IQI 2)
    - Abdominal aortic aneurysm repair (IQI 4)
    - Coronary artery bypass graft (IQI 5)
    - Percutaneous transluminal coronary angioplasty (IQI 6)
    - Carotid endarterectomy (IQI 7)

2 Statistical Methods

This section provides a brief overview of the structure of the administrative data from the Nationwide Inpatient Sample, and the statistical models and tools currently being used within the AHRQ Quality Indicators Project. We then propose several alternative statistical models and methods for consideration, including (1) models that account for trends in the response variable over time; and (2) statistical approaches that adjust for the potential positive correlation on patient outcomes from the same provider. We provide an overview of how these proposed alternative statistical approaches will impact the fitting of risk-adjusted models to the reference population, and on the tools that are provided to users of the QI methodology.

This is followed by an overview of the statistical modeling investigation, including (1) the selection of five IQIs to investigate in this report, (2) fitting current and alternative statistical models to data from the Nationwide Inpatient Sample, (3) statistical methods to compare parameter estimates between current and alternative modeling approaches using a Wald test-statistic, and (4) statistical methods to compare differences between current and alternative modeling approaches on provider-level model predictions (expected and risk-adjusted rates).

2.1 Structure of the Administrative Data

Hospital administrative data are collected as a routine step in the delivery of hospital services throughout the U.S., and provide information on diagnoses, procedures, age, gender, admission source, and discharge status on all admitted patients. These data can be used to describe the quality of medical care within individual providers (hospitals), within groups of providers (e.g., states, regions), and across the nation as a whole. Although in certain circumstances quality assessments based on administrative data are potentially prone to bias compared to possibly more clinically detailed data sources such as medical chart records, the fact that administrative data are universally available among the 37 States participating in the Healthcare Cost and Utilization Project (HCUP) allowed AHRQ to develop analytical methodologies to identify potential quality problems and success stories that merit further investigation and study.

The investigation in this report focuses on five select inpatient quality indicators, as applied to the Nationwide Inpatient Sample (NIS) from 2001-2003. The Nationwide Inpatient Sample represents a sample of administrative records from a sample of approximately 20 percent of the providers participating in the HCUP. There is significant overlap in the HCUP hospitals selected in the NIS, with several of the hospitals being repeatedly sampled in more than one year.

The NIS data is collected at the patient admission level. For each hospital admission, data is collected on patient age, gender, admission source, diagnoses, procedures, and discharge status. There is no unique patient identifier, so the same patient may be

represented more than once in the NIS data (with some patients potentially being represented more than once within the same hospital, and other patients potentially being represented more than once within multiple hospitals).

The purpose of the QI statistical models is to provide parameter estimates for each quality indicator that are adjusted for age, gender, and all patient refined diagnosis related group (APR-DRG). The APR-DRG classification methodology was developed by 3M, and provides a basis to adjust the QIs for the severity of illness or risk of mortality, and is explained elsewhere.

For each selected quality indicator, the administrative data is coded to indicate whether they contain the outcome of interest as follows:

> Let Y_{ijk} represent the outcome for the j^{th} patient admission within the i^{th} hospital, for the k^{th} Quality Indicator. Y_{ijk} is equal to one for patients who experience the adverse event, zero for patients captured within the appropriate reference population but do not experience the adverse event, and is missing for all patients that are excluded from the reference population for the k^{th} Quality Indicator.

For each Quality Indicator, patients with a missing value for Y_{ijk} are excluded from the analysis dataset. For all patients with $Y_{ijk} = 0$ or 1, appropriate age-by-gender and APR-DRG explanatory variables are constructed for use in the statistical models.

2.2 Current Statistical Models and Tools

The following two subsections provide a brief overview of the statistical models that are currently fit to the HCUP reference population, and the manner in which these models are utilized in software tools provided by the AHRQ Quality Indicators Project.

2.2.1 Models for the Reference Population

Currently, a simple logistic regression model is applied to three years of administrative data from the HCUP for each Quality Indicator, as follows:

$$logit(Pr(Y_{ijk} = 1)) = \beta_{k0} + \sum_{p=1}^{P_k} \alpha_{kp} \cdot (Age/Gender_p)_{ij} + \sum_{q=1}^{Q_k} \theta_{kq} \cdot (APR-DRG_q)_{ijk}, \quad (1)$$

where Y_{ijk} represents the response variable for the j^{th} patient in the i^{th} hospital for the k^{th} quality indicator; $(Age/Gender_p)_{ij}$ represents the p^{th} age-by-gender zero/one indicator variable associated with the j^{th} patient in the i^{th} hospital; and $(APR\text{-}DRG_q)_{ijk}$ represents the q^{th} APR-DRG zero/one indicator variable associated with the j^{th} patient in the i^{th} hospital for the k^{th} quality indicator.

For the k^{th} quality indicator, we assume that there are P_k age-by-gender categories and Q_k APR-DRG categories that will enter the model for risk-adjustment purposes.

The α_{kp} parameters capture the effects of each of the P_k age-by-gender categories on the QI response variable; and similarly, the θ_{kq} parameters capture the effects of each of the

Q_k APR-DRG categories on the QI response variable. The α_{kp} and θ_{kq} parameters each have ln(odds-ratio) interpretation, when compared to the reference population. The logit-risk of an adverse outcome for the reference population is captured by the β_{k0} intercept term in the model associated with the k^{th} Quality Indicator.

Model (1) can be fit using several procedures in SAS. For simplicity and consistency with other modeling approaches investigated in this report, we used SAS Proc Genmod to fit Model (1) to data from the Nationwide Inpatient Sample.

2.2.2 Software Tools Provided to Users

The AHRQ Quality Indicators Project provides access to software that can be downloaded by users to calculate expected and risk-adjusted QIs for their own sample of administrative data. The expected rate represents the rate that the provider would have experienced if it's quality of performance was identical to the reference (National) population, given the provider's actual case mix (e.g. age, gender, DRG and comorbidity categories). Expected rates are calculated based on combining the regression coefficients from the reference model (based on fitting Model (1) above to the reference HCUP population) with the patient characteristics from a specific provider.

Risk-adjusted rates are the estimated performance if the provider had an "average" patient mix, given their actual performance. It is the most appropriate rate upon which to compare *across hospitals*, and is calculated by adjusting the observed National Average Rate for the ratio of observed vs. expected rates at the provider-level:

$$\textit{Risk-adjusted rate} = (\textit{Observed Rate} \;/\; \textit{Expected Rate}) \times \textit{National Average Rate} \qquad (2)$$

The AHRQ Inpatient Quality Indicator software appropriately applies the National Model Regression Coefficients to the provider specific administrative records being analyzed to calculate both expected and risk-adjusted rates.

2.3 Alternative Statistical Methods

In the following sections, we propose several alternative statistical models and methods for consideration, including (1) models that account for trends in the response variable over time; and (2) statistical approaches that adjust for the potential positive correlation on patient outcomes from the same provider.

2.3.1 Adjusting for Trends over Time

The following alternative model formulation is proposed as a simple method for adjusting for the effects of quality improvement over time with the addition of a single covariate to Model (1):

$$\begin{aligned} logit(Pr(Y_{ijk}=1)) = \quad & \beta_{k0} + \sum_{p=1}^{P_k} \alpha_{kp} \cdot (Age/Gender_p)_{ij} + \sum_{q=1}^{Q_k} \theta_{kq} \cdot (APR-DRG_q)_{ijk} \\ & + \lambda_k \cdot (Year_{ijk} - 2002) \end{aligned} \qquad (3)$$

The parameter λ_k adjusts the model for a simple linear trend over time (on the logit-scale for risk of an adverse event), with the covariate ($Year_{ijk}$-2002) being a continuous variable that captures the calendar year that the j^{th} patient was admitted to the i^{th} hospital. This time-trend covariate is centered on calendar year 2002 in our analyses, to preserve a similar interpretation of the β_{k0} intercept term in Model (1), as our national reference dataset represents administrative records reported in calendar years 2001 through 2003.

Additional complexities can be introduced into the above simple time-trend model to investigate (1) non-linear time-trends on the logit scale, and (2) any changes over time in the age-by-gender or APR-DRG variable effects on risk of adverse outcomes. Such investigations were not explored within this report – but could be the subject of later data analyses. The authors of this report also suggest combining data over a longer period of time (e.g., five years or more) to better capture long-term trends in hospital quality of care.

The introduction of a time-trend into the model serves three purposes. First, it provides AHRQ (and users) with an understanding of how hospital quality is changing over time through the interpretation of the λ_k parameter (or similar time-trend parameters in any expanded time-trend model). Secondly, if the λ_k parameter is found to be statistically significant, the time-trend model will likely offer more precise expected and risk-adjusted rates. Thirdly, it may allow more accurate model predictions (expected and risk-adjusted rates for providers) when users apply a model based on older data to more recent data (often, a user might utilize software that is based on a 2001-2003 reference population to calculate rates for provider-specific data from calendar year 2005).

It is important to note that the authors of this report suggest exercising caution in extrapolating the results of the AHRQ models for prediction beyond the temporal range of observed data. However, it is our understanding that this is a common practice among users of the Quality Indicators Methodology. Given this type of use, a model which accounts for trends over time will likely provide more accurate predictions than a model that does not account for temporal trends.

2.3.2 Adjusting for Within-Provider Correlation

The current simple logistic regression modeling approach being used by AHRQ in the risk-adjusted model fitting assumes that all patient responses are independent and identically distributed. However, it is likely that responses of patients from within the same hospital may be correlated, even after adjusting for the effects of age, gender, severity of illness and risk of mortality. This anticipated positive correlation results from the fact that each hospital has a unique mixture of staff, policies and medical culture that combine to influence patient results. It is often the case that fitting simple models to correlated data results in similar parameter estimates, but biased standard errors of those parameter estimates – however, this does not always hold true. In the following two subsections, we provide an overview of generalized estimating equations (GEE) and generalized linear mixed modeling (GLIMMIX) approaches for adjusting the QI statistical models for the anticipated effects of within-provider correlation. These approaches will be investigated on a sample of five selected Quality Indicators to

determine whether (or not) the parameter estimates from a simple logistic regression model result in different parameter estimates or provider-level model predictions (expected and risk adjusted rates), in comparison to GEE or GLIMMIX approaches that account for the within-provider correlation.

2.3.2.1 Generalized Estimating Equations

The GEE methodology, introduced by Liang and Zeger (1986), provides a method of analyzing correlated data under the conceptual framework of Generalized Linear Models making use of Quasi-Likelihood theory under a marginal model for estimating the fixed effects portion of the model. The responses from studies with correlated data can often be organized into clusters, where observations from within a cluster may be statistically dependent, and observations from two different clusters are assumed independent. In the context of the AHRQ Quality Indicators project, the providers (hospitals) serve as clusters.

The marginal model for correlated binary outcomes (such as those from the AHRQ QI Project) can be thought of as a simple extension to a simple logistic regression model that directly incorporates the within-cluster correlation among patient responses from within the same hospital. To estimate the regression parameters in a marginal model, we make assumptions about the marginal distribution of the response variable (e.g. assumptions about the mean, it's dependence on the explanatory variables, the variance and the covariance among responses from within the same hospital). The cross-sectional model (Model (1)) and time-trend model (Model (3)) can be fit using the generalized estimating equations approach using SAS Proc Genmod, through the introduction of a repeated statement that accounts for the within-provider clustering. Appendix A, section A-1 provides additional detail for the GEE methodology.

2.3.2.2 Generalized Linear Mixed Models

In the previous section, we described marginal models for correlated/clustered data using a generalized estimating equations approach. An alternative approach for accounting for the within-hospital correlation is through the introduction of random effects into Model(1) as follows:

$$logit(Pr(Y_{ijk}=1)) = \beta_{k0} + \sum_{p=1}^{P_k} \alpha_{kp} \cdot (Age/Gender_p)_{ij} + \sum_{q=1}^{Q_k} \theta_{kq} \cdot (APR-DRG_q)_{ijk} + \gamma_{ki}, \quad (4)$$

where γ_{ki} is a random effect associated with each provider, and is assumed to follow a normal distribution with mean zero, and variance σ^2_{Hosp}. The time-trend model can be similarly expanded using a random effects model, as follows:

$$logit(Pr(Y_{ijk}=1)) = \beta_{k0} + \sum_{p=1}^{P_k} \alpha_{kp} \cdot (Age/Gender_p)_{ij} + \sum_{q=1}^{Q_k} \theta_{kq} \cdot (APR-DRG_q)_{ijk} + \lambda_k \cdot (Year_{ijk} - 2002) + \gamma_{0ki} + \gamma_{1ki} \cdot (Year_{ijk} - 2002) \quad (5)$$

where γ_{0ki} and γ_{1ki} are random intercept and slope terms associated with each provider (thus allowing each provider to depart from the fixed effects portion of the model with a provider-specific trend over time). In Model (5), we assume that γ_{0ki} and γ_{1ki} jointly follow a multivariate normal distribution with mean zero and covariance matrix $\Sigma = \begin{bmatrix} \sigma^2_{Hosp} & \sigma^2_{Hosp/Year} \\ \sigma^2_{Hosp/Year} & \sigma^2_{Year} \end{bmatrix}$.

Models (4) and (5) can be fit using SAS Proc GLIMMIX, and can also be expanded to allow for different probability distributions for the random effects (i.e. we can relax the assumption of normality for the random effects, if necessary).

2.3.3 Impact of Adopting Alternative Methods on Model Fitting and Tools

Currently, the risk-adjusted models for each quality indicator are fit to three calendar years of administrative data from the HCUP using various different data manipulations and model fitting procedures available from within the SAS software system. The addition of a time-trend covariate (or series of covariates) will not introduce any significant additional complexity to fitting these models to the reference (national) data. Adjusting the models for the anticipated positive correlation among patient responses from within the same hospital will require the use of Generalized Estimating Equations (GEE) approaches available through Proc GENMOD in SAS, or the use of Generalized Linear Mixed Modeling (GLIMMIX) approaches available through Proc GLIMMIX in SAS. Both of these methods are more computationally intense compared to fitting a simple logistic regression model, and may be subject to convergence problems and model mis-specification that is typical of such iterative modeling approaches.

Once the models are fit to the reference (national) population, integration of the modeling results into the software tools provided to users should be relatively straightforward. The introduction of a time-trend model would require the user to keep track of the calendar year associated with each patient response, for inclusion as a predictor variable in the model. If additional time-trend variables (either non-linear, or interactions with the other predictor variables) are introduced, the software can be quickly updated to accommodate these model changes.

Use of the GLIMMIX approach may yield additional information, in which the vector of random effects from the National Model can be exploited to determine the distribution of expected risk of adverse events (after adjusting for age, gender, severity of illness, and risk of mortality) across participating hospitals. This distribution can be used to identify (approximately) where within the national distribution of providers a particular hospital lies (currently, the AHRQ methodology only provides information related to whether an individual user is above or below the national mean). This use of the estimated vector of

random effects would require additional software development at AHRQ, as well as additional work to ensure that the GLIMMIX random effects model is adequately fitting the data from the reference population.

2.4 Overview of Statistical Modeling Investigation

The purpose of this report is to investigate the use of alternative modeling approaches to potentially adjust the risk-adjusted Quality Indicator Models for the effects of trends over time and the effects of positive correlation among responses from within the same hospital. In the following sections, we provide:

- an overview of the five Inpatient Quality Indicators that were selected for this investigation;
- a description of the various models fit to the Nationwide Inpatient Sample;
- statistical methodology used to assess whether or not the alternative modeling approaches yield parameter estimates that are significantly different from each other;
- statistical methodology used to assess whether or not the alternative modeling approaches yield provider-level estimates (expected and risk-adjusted rates of adverse events) that are significantly different from each other; and
- statistical methodology used for bootstrap sampling of the NIS sample to further assess differences in results attributable to model selection.

2.4.1 Selection of IQIs to Investigate

The following five Inpatient Quality Indicators were selected for exploration in this report (*with the descriptions for each QI copied directly from the AHRQ Guide to Prevention Quality Indicators*):

IQI 11: Abdominal Aortic Aneurism Repair Mortality Rate
Abdominal aortic aneurysm (AAA) repair is a relatively rare procedure that requires proficiency with the use of complex equipment; and technical errors may lead to clinically significant complications, such as arrhythmias, acute myocardial infarction, colonic ischemia, and death. The adverse event for this Quality Indicator is recorded as positive for any patient who dies with a code of AAA repair in any procedure field, and a diagnosis of AAA in any field. The reference population for this Quality Indicator includes any patient discharge with ICD-9-CM codes of 3834, 3844, and 3864 in any procedure field and a diagnosis code of AAA in any field. The reference population excludes patients with missing discharge disposition, who transfer to another short-term hospital, MDC 14 (pregnancy, childbirth, and puerperium), and MDC 15 (newborns and other neonates).

IQI 14: Hip Replacement Mortality Rate
Total hip arthroplasty (without hip fracture) is an elective procedure preformed to improve function and relieve pain among patients with chronic osteoarthritis, rheumatoid arthritis, or other degenerative processes involving the hip joint. The adverse event for this Quality Indicator is recorded as positive for any patient who dies with a code of paritial or full hip replacement in any procedure field. The reference population for this Quality Indicator includes any patient with procedure code of partial or full hip replacement in any field, and includes only discharges with uncomplicated cases: diagnosis codes for osteoarthritis of hip in any field. The reference population excludes patients with missing discharge disposition, who transfer to another short-term hospital, MDC 14 (pregnancy, childbirth, and puerperium), and MDC 15 (newborns and other neonates).

IQI 17: Acute Stroke Mortality Rate
Quality treatment for acute stroke must be timely and efficient to prevent potentially fatal brain tissue death, and patients may not present until after the fragile window of time has passed. The adverse event for this Quality Indicator is recorded as positive for any patient who dies with a principal diagnosis code of stroke. The reference population for this Quality Indicator includes any patient aged 18 or older with a principal diagnosis code of stroke. The reference population excludes patients with missing discharge disposition, who transfer to another short-term hospital, MDC 14 (pregnancy, childbirth, and puerperium), and MDC 15 (newborns and other neonates).

IQI 19: Hip Fracture Mortality Rate
Hip fractures, which are a common cause of morbidity and functional decline among elderly patients are associated with a significant increase in the subsequent risk of mortality. The adverse event for this Quality Indicator is recorded as positive for any

patient who dies with a principal diagnosis code of hip fracture. The reference population for this Quality Indicator includes any patient aged 18 or older with a principal diagnosis code of hip fracture. The reference population excludes patients with missing discharge disposition, who transfer to another short-term hospital, MDC 14 (pregnancy, childbirth, and puerperium), and MDC 15 (newborns and other neonates).

IQI 25: Bilateral Cardiac Catherization Rate
Righ- side coronary catheterization incidental to left side catheterization has little additional benefit for patient without clinical indications for right-side catheterization. The adverse event for this Quality Indicator is recorded as positive for any patient with coronary artery disease who has simultaneous right and left heart catheterizations in any procedure field (excluding valid indications for right-sided catherization in any diagnosis field, MDC 14 (pregnancy, childbirth, and puerperium), and MDC 15 (newborns and other neonates)). The reference population for this Quality Indicator includes any patient with coronary artery disease discharged with heart catheterization in any procedure field. The reference population excludes patients with MDC 14 (pregnancy, childbirth, and puerperium), and MDC 15 (newborns and other neonates).

2.4.2 Fitting Current and Alternative Models to NIS Data

For each selected IQI, we fit risk-adjusted cross-sectional and time-trend adjusted models using a simple logistic regression, generalized estimating equations and generalized linear mixed modeling approach (for a combined six models for each IQI). The models were fit to three-years of combined data from the Nationwide Inpatient Sample, which represents an approximate 20 percent sample of hospitals from within the HCUP (with administrative records included for all patients treated within each selected hospital).

The data were processed to eliminate any patient records that are excluded from the reference population prior to modeling (thus, only patients with a zero or one response were included in the analysis for each IQI). The form of the model followed what was included in the models currently fit to the HCUP data – with minor modifications to remove covariates that represented very sparse cells.

For the GEE and GLIMMIX approaches, we retained both the robust and model-based variance/covariance matrices for the vector of parameter estimates, to allow for appropriate statistical comparisons using both methods. We also retained the vector of random effects from the GLIMMIX approach, to assess for distributional assumptions.

2.4.3 Fitting Current and Alternative Models to Boot Strap Samples of NIS Data

Bootstrapping is a statistical method used for estimating statistical modeling error based on resampling, with the resulting error estimates often being used for choosing among various models. Large sample theory suggests that the parameter estimates from the three modeling approaches: Simple Logistic, GEE and Generalized linear models should converge to the same parameter estimates ($\hat{\beta}$). Standard errors of $\hat{\beta}$ will likely diverge between these three approaches due to manner in which each method accounts for the

within-hospital correlation. These differences in standard errors would affect confidence intervals for the AHRQ QIs.

The bootstrap analyses used the NIS sample data as a population from which repeated samples were drawn. To assess whether the parameter estimates from the three modeling approaches converge to the same values as a function of sample size, we investigated bootstrap samples of varying sizes – expressed as a proportion of the size of the NIS sample itself. Our analyses focused on bootstrap samples that had a total patient population of approximately 25%, 50%, 100%, 150%, 250% and 500% the size of the NIS sample. For each target population size described above, 100 bootstrap samples were selected and evaluated to compare the three statistical modeling methods.

The bootstrap samples were conducted at the hospital level, in order to preserve the within-hospital correlation observed within the NIS sample. For each QI, the observed NIS sample was divided into four strata based on the summary of patient counts within each hospital (with the summary counts representing the number of patients in the denominator for each QI). Each strata contained an approximately equal number of patients, however the first strata consisted of the largest hospitals (and therefore had fewer hospitals represented in the strata), the second strata consisted of the next largest hospitals, with the fourth strata consisting of the smallest hospitals (thereby having the largest number of hospitals represented in this strata).

Bootstrap sampling occurred with selection probabilities weighted by the number of patients represented within each hospital (i.e. proportional to size sampling). Due to the fact that sampling was conducted at the hospital level within each strata, the sampling process was conducted sequentially with replacement until the target sample size of patients was exceeded within each strata. We then determined whether the sample size was closer to the target by including/excluding the last hospital selected in the sampling scheme.

As each hospital was selected into the bootstrap sample, it was given a unique hospital identifier for use in the statistical models, and all patient records (across all years) within the selected hospital were utilized in the analysis. The data were then combined across the four strata, and evaluated using the three statistical modeling techniques (simple logistic regression, generalized estimating equations, and generalized linear mixed models) for the original cross sectional model and the model that adjusts for trends over time.

This bootstrap sampling and model fitting was repeated 100 times at each 25, 50, 100, 150, 250, and 500 percent of the observed NIS sample size. The results of these bootstrap analyses were evaluated using methods described below in Section 2.5.

2.5 Methods to Compare Parameter Estimates

The current and the two alternative modeling approaches were investigated on a sample of five selected Quality Indicators to determine whether (or not) the parameter estimates

from a simple logistic regression model result in different parameter estimates or provider-level model predictions (expected and risk adjusted rates), in comparison to GEE or GLIMMIX approaches that account for the within-provider correlation. The following three subsections we provide statistical tests to assess for significant differences between a specific pair of modeling approaches.

When comparing two or more sets of results, researchers naturally focus on pairwise differences. For example, does the GEE method consistently provide lower estimates of the intercept relative to the simple logistic regression model?

The answers to these questions may be expressed as either absolute or relative differences based on the modeling results. An absolute difference is a subtraction; a relative difference is a ratio.

In the methods described in the subsections below, differences between the parameter estimates from any specific pair of models were calculated and used for (1) a Global Wald Test, and (2) paired ttest for testing intercept differences, as follows:

$$Difference(SLR - GEE) = \left(\hat{\beta}_{SLR} - \hat{\beta}_{GEE}\right)$$

Statistics presenting relative difference relate to the average between the two methods being compared:

$$\text{Re}\,lativeDifference(SLR - GEE) = \frac{Abs\left(\hat{\beta}_{SLR} - \hat{\beta}_{GEE}\right)}{\frac{1}{2}\left[Abs\left(\hat{\beta}_{SLR}\right) + Abs\left(\hat{\beta}_{GEE}\right)\right]}$$

In the above formulas, $\hat{\beta}_{SLR}$ and $\hat{\beta}_{GEE}$ represent the parameter estimates from the simple logistic and GEE models. The relative differences between (1) simple logistic and Generalized linear models and (2) GEE and Generalized linear models were calculated in a similar fashion.

In the sections that follow describing the results of the bootstrap sampling investigations, measures of pairwise modeling differences and relative differences (i.e. SLR vs GEE, SLR vs GLIMMIX, and GEE vs GLIMMIX) are provided based on the mean vector of parameter estimates across 100 bootstrap samples at a particular target sample size (relative to the size of the NIS sample).

2.5.1 Box Plots

Box plots (Chambers 1983) are an excellent tool for conveying location and variation information in data sets, particularly for detecting and illustrating location and variation changes between different groups of data. A box plot displays the median (represented

by the center horizontal line), the 25th percentile (represented by the bottom of the box), and the 75th percentile (represented by the top of the box). The vertical lines, or whiskers, are drawn from the box to the most extreme point within 1.5 * interquartile range. (An interquartile range is the distance between the 25th and the 75th percentiles.) Any value more extreme than this is identified individually with stars.

Box plots of both specific parameter estimates from the three modeling approaches, as well as differences in parameter estimates between any pair of specific models were produced for each of the five selected QIs to graphically compare the estimates from three different modeling approaches.

2.5.2 Wald Test

Given that the parameter estimates from each of the logistic regression models follow an approximate normal distribution (as shown in McCulloch & Nelder, 1989 within the Generalized Linear Model conceptual framework), a Wald Statistic can be used to assess whether there are statistically significant differences between the full set of parameter estimates yielded from a specific pair of modeling approaches. For example, when comparing the simple linear regression model results to the results of the GEE approach, we have the following Wald Statistics:

$$W_1 = \left(\hat{\beta}_{SLR} - \hat{\beta}_{GEE}\right)^T V_{SLR/GEE}^{-1} \left(\hat{\beta}_{SLR} - \hat{\beta}_{GEE}\right)$$

In the above formulas, $\hat{\beta}_{SLR}$ and $\hat{\beta}_{GEE}$ represent the parameter estimates from the simple logistic and GEE models and $V_{SlR/GEE}^{-1}$ represents an appropriate covariance matrix for the difference between regression parameters from the simple logistic and GEE models.

Under the null hypothesis, that there are no statistically significant differences between the simple logistic regression model and GEE parameter estimates, the above three Wald test statistics are expected to follow a $X_{(p)}^2$ distribution (a Chi-squared distribution with p degrees of freedom, where p represents the number of explanatory variables that were used within the statistical model).

The Global Wald test described above could be extended to asses the differences in parameter estimates between (1) simple logistic and generalized linear models (SLR vs. GLIMMIX) and (2) GEE and generalized linear models (GEE vs. GLIMMIX).

The results provided later in the report focus on the application of the Wald Test in two ways:

1. We applied the Wald Test Statistic to each pair of parameter estimates as fit to the NIS sample. In this application, we did not have an appropriate measure of $V_{SlR/GEE}^{-1}$ - rather, we have the inverse covariance matrix of parameter estimates from the SLR model fit to the NIS sample (V_{SlR}^{-1}), and we have the inverse

covariance matrix from the GEE model fit to the NIS sample (V_{GEE}^{-1}). Thus, by substituting in either of these inverse covariance matrices – the Wald Test Statistic answers the question: When fitting a model to the NIS sample – did the use of the GEE method provide parameter estimates that were statistically different from the SLR method (assuming that the SLR method is the valid method, and using (V_{SlR}^{-1}) as the inverse variance estimate)? Alternatively, we could address the parallel question: When fitting a model to the NIS sample – did the use of the SLR method provide parameter estimates that were statistically different from the GEE method (assuming that the GEE method is the valid method, and using (V_{GEE}^{-1}) as the inverse variance estimate)?

2. We also applied the Wald Statistic across each set of bootstrap results. Under the generalized linear model framework, we can assume that the 100 observed vectors of pairwise differences in parameter estimates follow a multivariate normal distribution. Thus a simple mean vector of parameter estimate differences and corresponding covariance matrix can be constructed and used as input for a Wald Test to appropriately assess for pairwise modeling differences across all parameters fit in the model.

2.5.3 T-Tests

One-sample T-test (or paired t-test) can be used to test whether there are statistically significant differences in estimates of a single parameter (e.g., the intercept) between a specific pair of modeling approaches. Since each pair of models fit to identical data sets, we assume a natural pairing of the estimates exist and utilizing the correlation among pairs of estimates from models fit to identical datasets will result in higher power to detect existing differences between the means. Here the null hypothesis is to test whether the mean change in intercept estimates from any two methodologies are significantly different from zero i.e. Ho: $\mu_1 - \mu_2 = 0$. Note that this test of hypothesis assumes that the differences between paired observations are normally distributed. The test statistic that was used to make a decision whether or not to reject the null hypothesis is given by the following formula.

$$T = \frac{\overline{D}}{S\sqrt{\frac{1}{n}}}$$

where $\overline{D}$ is the sample mean of the paired differences and is the S^2 is the sample variance of the paired differences, n is the number of paired observations and T is the student-t quantile with n-1degrees of freedom under the null hypothesis.

The paired t-test described above could be extended to each of the parameters included in the model, but only differences in intercept estimates between (1) simple logistic and generalized linear models (SLR vs. GLIMMIX) and (2) GEE and generalized linear models (GEE vs. GLIMMIX) using t-test are presented in this report.

2.5.4 Methods to Compare Provider-Level Model Predictions

For each select IQI, we identified a simple random sample of 50 providers to use for assessing differences between provider-level model predictions (both expected and risk-adjusted rates). Simple descriptive statistics (mean and standard deviation) were generated for the distribution differences in provider-level model predictions to assess whether (or not) changes in the model might result in any potential bias or increased variability in provider-level estimates.

The distributional summaries were conducted separately for the cross-sectional and time-trend models (so that the statistics isolate any differences attributable to adjusting the models for the potential correlation among responses within the same provider).

Subsequent analyses will be conducted at a later date to provide comparisons between the cross-sectional and time-trend models within each model type (and potentially across model types).

3 Results

All models were successfully fit to the NIS data source. The GLIMMIX approach initially suffered from convergence problems while using the default optimization techniques, but converged for all five IQIs (both cross sectional and time-trend adjusted models) when using Newton-Raphson optimization with ridging. Due to the large sample size of the dataset, the personal computer used to fit the model ran out of memory when calculating the robust variance-covariance matrix associated with the parameter estimates for IQI-25 with the GLIMMIX approach (the computer had 2GB of RAM).

Section 3.1 below provides summary statistics for the quality indicator response variables that were modeled from within the National Inpatient Sample. Sections 3.2 through 3.6 provide model results for each of the five select IQIs explored in this report.

3.1 Summary Statistics for NIS Data

Table 3.1 below provides summary statistics for the five selected quality indicators. The summary statistics include:

- The number of adverse events observed
- The number of patients in the reference population
- The number of hospitals that had patients within the reference population
- The mean response (proportion of patients who experienced the adverse event)
- The standard error associated with the mean response
- Select percentiles from the distribution (5^{th}, 25^{th}, 50^{th}, 75^{th}, and 95^{th})

Separate summary statistics were generated for each year of data (2001, 2002, and 2003) and then for all years combined. Prior to calculating the mean, standard error, and percentiles, the responses were averaged at the hospital level. These statistics therefore represent the distribution of hospital mean responses, and are presented in two ways (weighted and unweighted). The weighted results weigh each provider according to the number of patients observed within the reference population, whereas the unweighted results treat each hospital equally.

The weighted analysis mean was used as the National Average Rate when constructing the provider-specific risk-adjusted rates. Conceptually, the standard-error of the mean from the unweighted analysis should be proportional to the standard-error of the mean from the vector of random effects intercepts generated using the GLIMMIX approach from the cross-sectional model (although we anticipate that the variability of the random effects would be smaller due to the fact that other factors (age, gender, severity of illness and risk of mortality) are explaining variability in the Quality Indicator response variable.

Table 3.1 Summary Statistics (at the Provider Level) for the Five Selected Quality Indicators

IQI	Year	Analysis Type	Summary Statistics									
			n_{cases}	n_{Pop}	n_{Hosp}	Mean	Std Error	5th %ile	25th %ile	50th %ile	75th %ile	95th %ile
11	2001	Unweighted	726	8833.0	470	0.141	0.010	0.000	0.000	0.067	0.161	0.500
		Weighted				0.082	0.004	0.000	0.032	0.063	0.107	0.222
	2002	Unweighted	655	8099.0	449	0.152	0.011	0.000	0.000	0.070	0.188	0.750
		Weighted				0.081	0.005	0.000	0.028	0.063	0.098	0.231
	2003	Unweighted	558	8144.0	447	0.132	0.011	0.000	0.000	0.050	0.143	0.500
		Weighted				0.069	0.004	0.000	0.027	0.048	0.078	0.200
	All Years	Unweighted	1939	25076	1050	0.149	0.007	0.000	0.000	0.071	0.176	0.667
		Weighted				0.077	0.003	0.000	0.034	0.057	0.094	0.208
14	2001	Unweighted	104	34628	687	0.004	0.001	0.000	0.000	0.000	0.000	0.022
		Weighted				0.003	0.000	0.000	0.000	0.000	0.001	0.015
	2002	Unweighted	112	39677	682	0.004	0.001	0.000	0.000	0.000	0.000	0.020
		Weighted				0.003	0.000	0.000	0.000	0.000	0.002	0.014
	2003	Unweighted	86	39068	689	0.004	0.001	0.000	0.000	0.000	0.000	0.016
		Weighted				0.002	0.000	0.000	0.000	0.000	0.000	0.012
	All Years	Unweighted	302	113373	1544	0.004	0.000	0.000	0.000	0.000	0.000	0.020
		Weighted				0.003	0.000	0.000	0.000	0.000	0.003	0.013
17	2001	Unweighted	12616	108886	957	0.115	0.003	0.000	0.065	0.105	0.143	0.250
		Weighted				0.116	0.002	0.045	0.086	0.115	0.140	0.186
	2002	Unweighted	12298	109670	959	0.113	0.003	0.000	0.067	0.103	0.143	0.250
		Weighted				0.112	0.001	0.050	0.084	0.112	0.134	0.183
	2003	Unweighted	12385	108720	951	0.108	0.003	0.000	0.060	0.100	0.139	0.238
		Weighted				0.114	0.001	0.046	0.087	0.112	0.137	0.194
	All Years	Unweighted	37299	327276	2095	0.111	0.002	0.000	0.069	0.102	0.141	0.238
		Weighted				0.114	0.001	0.053	0.089	0.112	0.136	0.184
19	2001	Unweighted	1916	60318	809	0.036	0.003	0.000	0.000	0.026	0.044	0.095
		Weighted				0.032	0.001	0.000	0.017	0.031	0.043	0.066
	2002	Unweighted	2008	60597	800	0.039	0.003	0.000	0.000	0.026	0.047	0.092
		Weighted				0.033	0.001	0.000	0.019	0.031	0.044	0.069
	2003	Unweighted	1919	60559	804	0.037	0.003	0.000	0.000	0.024	0.043	0.100
		Weighted				0.032	0.001	0.000	0.019	0.029	0.040	0.070
	All Years	Unweighted	5843	181474	1806	0.038	0.002	0.000	0.000	0.027	0.044	0.091
		Weighted				0.032	0.000	0.005	0.020	0.031	0.041	0.065
25	2001	Unweighted	22858	283099	437	0.098	0.006	0.000	0.032	0.065	0.121	0.293
		Weighted				0.081	0.003	0.018	0.037	0.061	0.103	0.208
	2002	Unweighted	21455	286259	444	0.085	0.005	0.000	0.030	0.057	0.111	0.255
		Weighted				0.075	0.003	0.014	0.038	0.059	0.097	0.196
	2003	Unweighted	21697	298744	455	0.086	0.005	0.000	0.026	0.056	0.103	0.267
		Weighted				0.073	0.003	0.015	0.031	0.056	0.087	0.201
	All Years	Unweighted	66010	868102	997	0.090	0.004	0.000	0.029	0.057	0.107	0.261
		Weighted				0.076	0.002	0.017	0.037	0.057	0.095	0.208

3.2 IQI 11 – Abdominal Aortic Aneurysm Repair Mortality Rate

3.2.1 Model Parameter Estimates from Fitting Models to NIS Sample Data

Below we provide the model parameter estimates from fitting the simple logistic regression, generalized estimating equations, and generalized linear mixed model to the 2001-2003 Nationwide Inpatient Sample for IQI 11 (Abdominal Aortic Aneurysm Repair Mortality Rate). Table 3.2.1a provides the parameter estimates associated with the cross-sectional model, and Table 3.2.1b provides the parameter estimates associate with the model that adjusts for a simple linear trend over time. Across these two tables, we see that the parameter estimates and associated standard errors are quite comparable among the three modeling approaches. In fact, the estimated correlation coefficient from the GEE modeling approach is nearly zero in both the cross-sectional (ρ=0.0073) and time-trend adjusted (ρ=0.0073) models. The estimated variance components associated with provider-specific random effects from the GLIMMIX model were subtle, yet statistically significant in both models ($\sigma^2_{\text{Btw Hosp}} = 0.191$ for the cross sectional model, and 0.185 for the time-trend model). The variance component that captures provider-specific variation in the time-trend slope was estimated as zero.

Table 3.2.1c provides the Wald Statistics to determine whether there are statistically significant differences between the vector of parameter estimates generated by each modeling approach. The Wald Statistics consider pair-wise comparisons, and suggest that there were no significant differences between the different modeling approaches for IQI 11 when applied to the data from the NIS.

Table 3.2.1a Parameter Estimates from Cross Sectional Models fit to IQI-11 (Abdominal Aortic Artery Repair Mortality Rate)

Parameter	Simple Logistic Regression Model			Generalized Estimating Equations Model			Generalized Linear Mixed Model		
	Estimate	Std Err	p value	Estimate	Std Err	p value	Estimate	Std Err	p value
Intercept	-1.218	0.258	0.000	-1.198	0.255	0.000	-1.177	0.262	0.000
SEX	0.074	0.184	0.689	0.062	0.181	0.734	0.038	0.188	0.839
AGE1	-0.905	0.111	0.000	-0.898	0.109	0.000	-0.938	0.113	0.000
AGE13	-0.410	0.128	0.001	-0.408	0.125	0.001	-0.431	0.130	0.001
AGE15	0.354	0.199	0.076	0.354	0.196	0.071	0.393	0.203	0.052
AGE27	0.261	0.231	0.259	0.253	0.227	0.265	0.281	0.235	0.233
C2	0.626	1.147	0.585	0.684	1.118	0.541	0.750	1.164	0.519
C3	-0.697	1.063	0.512	-0.645	1.028	0.530	-0.555	1.069	0.604
C4	-0.507	1.075	0.637	-0.403	1.023	0.693	-0.354	1.089	0.745
C5	-3.141	0.299	0.000	-2.979	0.291	0.000	-3.099	0.302	0.000
C6	-2.619	0.272	0.000	-2.487	0.267	0.000	-2.565	0.275	0.000
C7	-0.938	0.254	0.000	-0.893	0.251	0.000	-0.880	0.257	0.001
C8	1.451	0.239	0.000	1.451	0.237	0.000	1.496	0.242	0.000
C9	-1.465	0.243	0.000	-1.366	0.241	0.000	-1.387	0.247	0.000
ρ				0.0073	.	.			
$\sigma^2_{\text{Btw Hosp}}$							0.191	0.043	.

The effect of the *YEAR* parameter (which captures the trend over time) was highly significant for all three modeling approaches, as seen in Table 3.2.1b below.

Table 3.2.1b Parameter Estimates from Time Trend Models fit to IQI-11 (Abdominal Aortic Artery Repair Mortality Rate)

Parameter	Simple Logistic Regression Model			Generalized Estimating Equations Model			Generalized Linear Mixed Model		
	Estimate	Std Err	p value	Estimate	Std Err	p value	Estimate	Std Err	p value
Intercept	-1.224	0.258	0.000	-1.203	0.255	0.000	-1.182	0.262	0.000
SEX	0.076	0.184	0.681	0.064	0.181	0.725	0.041	0.187	0.828
AGE1	-0.907	0.111	0.000	-0.900	0.109	0.000	-0.939	0.113	0.000
AGE13	-0.409	0.128	0.001	-0.407	0.125	0.001	-0.429	0.130	0.001
AGE15	0.355	0.199	0.075	0.355	0.196	0.069	0.394	0.203	0.052
AGE27	0.261	0.232	0.260	0.252	0.227	0.267	0.279	0.235	0.235
C2	0.678	1.146	0.554	0.723	1.121	0.519	0.769	1.165	0.509
C3	-0.677	1.062	0.524	-0.629	1.029	0.541	-0.546	1.069	0.609
C4	-0.485	1.075	0.652	-0.381	1.022	0.710	-0.325	1.087	0.765
C5	-3.152	0.299	0.000	-2.990	0.291	0.000	-3.110	0.302	0.000
C6	-2.626	0.272	0.000	-2.496	0.267	0.000	-2.573	0.275	0.000
C7	-0.943	0.254	0.000	-0.897	0.252	0.000	-0.885	0.258	0.001
C8	1.450	0.239	0.000	1.449	0.237	0.000	1.493	0.242	0.000
C9	-1.453	0.243	0.000	-1.357	0.241	0.000	-1.379	0.247	0.000
YEAR	-0.108	0.034	0.001	-0.105	0.035	0.002	-0.101	0.037	0.007
ρ				0.0071	.	.			
σ^2_{Hosp}							0.185	0.043	.
σ^2_{Year}							0.000	.	.

Table 3.2.1c Wald Test Statistics and (P-Value) Comparing Models fit to IQI-11 (Abdominal Aortic Artery Repair Mortality Rate) in the NIS Sample

	Cross Sectional Model			Time Trend Model		
	SLR	GEE	GLIMMIX	SLR	GEE	GLIMMIX
SLR		7.880 (0.895)	6.21 (0.961)		7.624 (0.938)	6.05 (0.979)
GEE	6.070 (0.965)		2.452 (1.000)	5.907 (0.981)		2.400 (1.000)
GLIMMIX	4.117 (0.995)	2.311 (1.000)		4.047 (0.998)	2.260 (1.000)	

- Wald Test uses the estimated covariance matrix from the model listed in each row

Table 3.2.1d below provides the mean and standard deviation of differences between model predictions (expected rates above the diagonal, and risk-adjusted rates below the diagonal) from a random sample of 50 providers within the NIS reference population for IQI 11. For example, the mean difference in expected rates between the simple logistic regression and GEE approaches for the cross sectional model was 0.004 (relative to a national mean response rate of 0.077 from Table 3.1). Mean differences (and standard deviations) attributable to model specification (simple logistic vs GEE vs GLIMMIX) for the risk-adjusted rates appear to be higher than the expected rates.

Table 3.2.1d Estimated Differences (and Standard Deviation) in Provider-Level Model Predictions of Expected and Risk Adjusted Rates for IQI-11 (Abdominal Aortic Artery Repair Mortality Rate)

	Cross Sectional Model			Time Trend Model		
	SLR	**GEE**	**GLIMMIX**	**SLR**	**GEE**	**GLIMMIX**
SLR		-0.004 (0.001)	-0.002 (0.002)		-0.004 (0.001)	-0.002 (0.002)
GEE	0.061 (0.065)		0.001 (0.002)	0.061 (0.068)		0.001 (0.002)
GLIMMIX	0.033 (0.032)	-0.028 (0.037)		0.034 (0.034)	-0.027 (0.038)	

***** In each 3x3 table above, Expected Rate Differences (and Standard Deviations) are above the diagonal, and Adjusted Rate Differences (and Standard Deviations) are below the diagonal.

3.2.2 Model Parameter Estimates from Fitting Models to Boot Strap Samples of NIS Data

Both cross sectional and time-trend adjusted models (especially GLIMMIX) fit for quality indicator IQI-11 (Abdominal Aortic Aneurysm Repair Mortality Rate) for bootstrap samples with target sample size of 25% to 100% of the observed NIS sample suffered greatly from rare occurrence of the response variable within selected hospitals – thereby producing degenerate solutions or suffering from convergence problems. Our bootstrap results therefore focus on bootstrap sample sizes that are approximately 150%, 250% and 500% larger than the NIS sample.

Tables 3.2.2a and 3.2.2b provide an overview of the results for the cross sectional and time-trend adjusted models fit to 100 bootstrap samples with population size approximately 150% of the size of the NIS sample. Each table provides the mean across the 100 bootstrap samples of the model parameter estimates and standard errors from the Simple Logistic Regression Model. The table also provides average absolute and relative differences between each pair of modeling approaches. Box plots of model specific intercept estimates and differences in intercept estimates between each pair of modeling approaches when target sample sizes ranging from 150% to 500% of the observed NIS samples for cross-sectional models are displayed in figures 3.2.2a and 3.2.2c. Figures 3.2.2b and 3.2.2d presents the similar boxplots for the time-trend models. The results (p-values) from global Wald tests performed to assess whether there are statistically significant difference between a pair of model parameter estimates are provided in Table 3.2.2c. Results (p-values) from pair-wise T-test to test for intercept differences are also included in Table 3.2.2c. Both Wald and T-tests are performed at 5% level of significance.

As seen from tables 3.2.2a and 3.2.2b parameter estimates from the three modeling approaches are very close for all the independent covariates included in the models. Average differences in parameter estimates between any pair of models are comparatively small. Relative differences between model estimates for most part are small except for some parameters like gender, C3, and C4. The variability in parameter

estimates and differences (intercept) are quite pronounced as seen in box plots. As the sample size increases the variability decreases as expected.

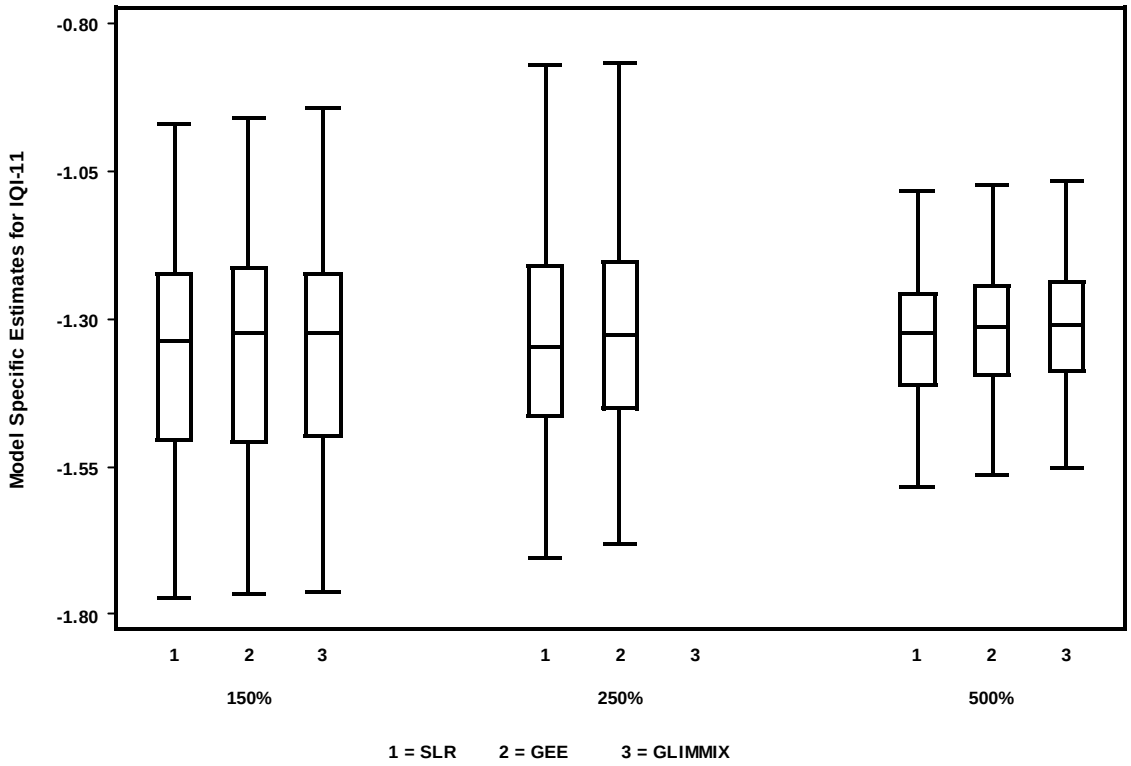
The Box plots of estimated intercepts for each modeling method in Figures 3.2.2a (cross sectional) and 3.2.2b (time-trend adjusted) demonstrate a high degree of overlap in the distribution of estimated intercepts among the three modeling approaches. However, Figures 3.2.2c and 3.2.2d show that the distribution of pairwise differences in parameter estimates from each individual bootstrap sample are not concentrated near the reference line at zero – indicating that there may be very subtle, yet statistically significant differences between the three modeling techniques. The Global Wald tests demonstrate highly significant differences between modeling techniques across all target sample sizes and modeling approach combinations. The paired t-test for intercept differences were not statistically significant for GEE vs. GLIMMIX for the cross-sectional model, and SLR vs. GEE and GEE vs. GLIM for time trend models at target sample size of 150%. Also there are no statistically significant differences between estimated intercepts between GEE and GLIMMIX models for target sample sizes at 250 and 500% of the NIS sample as seen from table 3.2.2c.

Table 3.2.2a. Average Models Estimates and Relative Standard Errors from Simple Logistic Model Fit, Differences and Relative Differences in Model Estimates from a Specific Pair of Modeling Approaches for the Inpatient Quality Indicator IQI-11 (Abdominal Aortic Artery Repair Mortality Rate) For Each Parameter in the Model (Cross-Sectional)

Cross-Sectional								
Parameter	Simple Logistic Regression		Difference between model estimates from a pair of modeling approaches			Relative Difference between model estimates from a pair of modeling approaches		
	Estimate	Std. Error	SLR vs. GEE	SLR vs. GLIM	GEE vs. GLIM	SLR vs. GEE	SLR vs. GLIM	GEE vs. GLIM
Intercept	-1.359	0.2264	-0.011	-0.012	-0.001	0.008	0.009	0.001
SEX	0.121	0.1558	0.012	0.038	0.026	0.101	0.372	0.273
AGE1	-0.862	0.0931	-0.003	0.035	0.038	0.004	0.039	0.043
AGE13	-0.377	0.1070	0.000	0.025	0.025	0.000	0.064	0.064
AGE15	0.300	0.1682	-0.002	-0.042	-0.039	0.008	0.130	0.122
AGE27	0.265	0.1943	0.008	-0.024	-0.032	0.030	0.086	0.116
C2	0.743	0.9635	-0.049	-0.114	-0.065	0.063	0.142	0.079
C3	-0.358	0.8428	-0.026	-0.114	-0.088	0.075	0.378	0.305
C4	-0.241	0.8506	-0.073	-0.118	-0.044	0.360	0.646	0.304
C5	-3.078	0.2592	-0.132	-0.053	0.080	0.044	0.017	0.027
C6	-2.620	0.2390	-0.112	-0.060	0.052	0.044	0.023	0.020
C7	-0.835	0.2227	-0.039	-0.065	-0.026	0.048	0.081	0.033
C8	1.516	0.2110	-0.005	-0.055	-0.050	0.003	0.036	0.032
C9	-1.426	0.2144	-0.092	-0.087	0.005	0.067	0.063	0.004

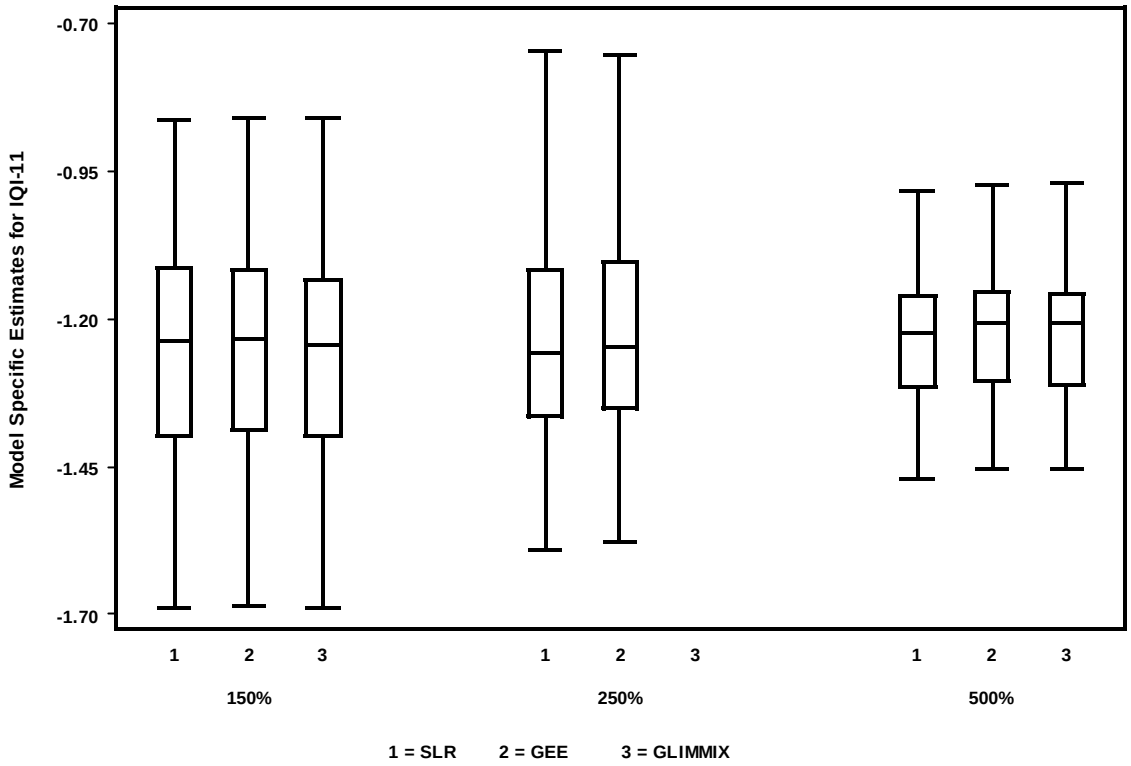
Table 3.2.2b. Average Models Estimates and Relative Standard Errors from Simple Logistic Model Fit, Differences and Relative Differences in Model Estimates from a Specific Pair of Modeling Approaches for the Inpatient Quality Indicator IQI-11 (Abdominal Aortic Artery Repair Mortality Rate) For Each Parameter in the Model (Cross-Sectional)

Time-Trend								
Parameter	Simple Logistic Regression		Difference between model estimates from a pair of modeling approaches			Relative Difference between model estimates from a pair of modeling approaches		
	Estimate	Std. Error	SLR vs. GEE	SLR vs. GLIM	GEE vs. GLIM	SLR vs. GEE	SLR vs. GLIM	GEE vs. GLIM
Intercept	-1.263	0.2281	-0.006	-0.002	0.004	0.005	0.002	0.003
SEX	0.125	0.1557	0.012	0.038	0.026	0.098	0.357	0.261
AGE1	-0.864	0.0932	-0.003	0.033	0.037	0.004	0.038	0.042
AGE13	-0.375	0.1071	0.000	0.025	0.024	0.001	0.064	0.063
AGE15	0.300	0.1681	-0.002	-0.041	-0.039	0.008	0.129	0.121
AGE27	0.263	0.1944	0.008	-0.023	-0.031	0.032	0.083	0.115
C2	0.781	0.9634	-0.039	-0.087	-0.049	0.048	0.106	0.058
C3	-0.353	0.8429	-0.025	-0.108	-0.083	0.074	0.361	0.289
C4	-0.215	0.8504	-0.072	-0.117	-0.045	0.401	0.745	0.372
C5	-3.092	0.2594	-0.132	-0.053	0.079	0.044	0.017	0.026
C6	-2.632	0.2391	-0.111	-0.060	0.051	0.043	0.023	0.020
C7	-0.843	0.2229	-0.040	-0.065	-0.025	0.048	0.080	0.032
C8	1.511	0.2111	-0.005	-0.053	-0.048	0.003	0.035	0.031
C9	-1.418	0.2145	-0.088	-0.082	0.006	0.064	0.060	0.005
Year	-0.098	0.0281	-0.004	-0.011	-0.006	0.046	0.118	0.071



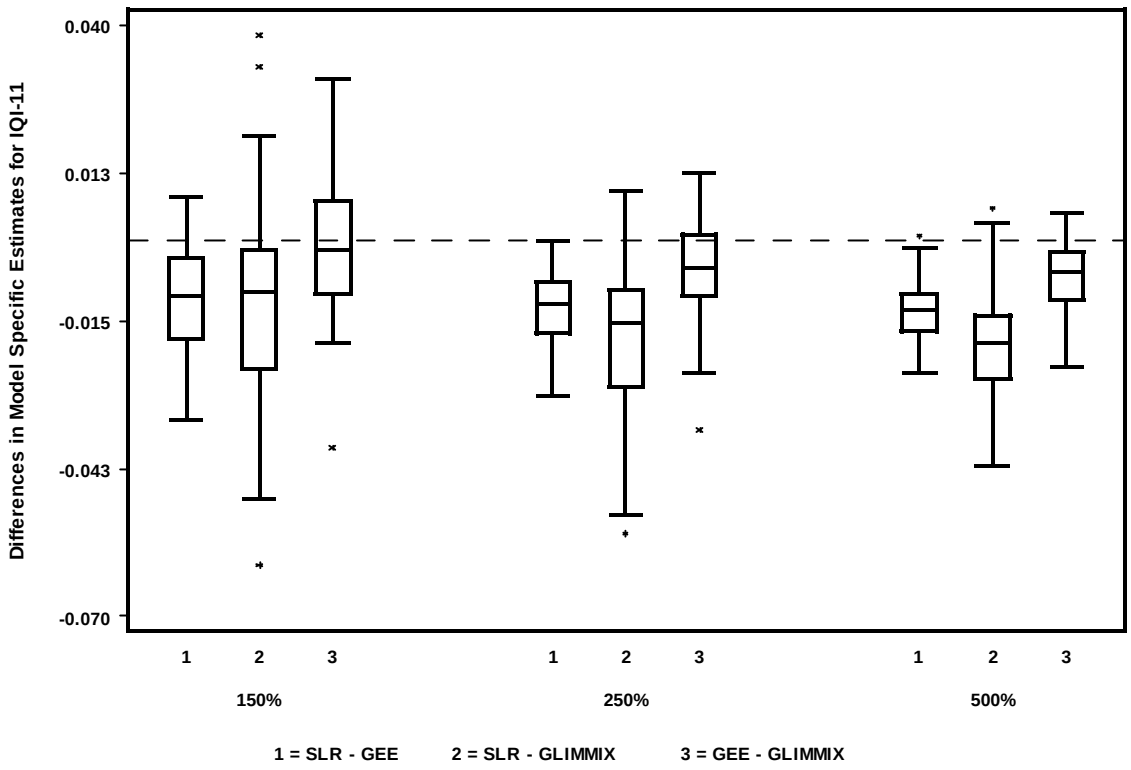
Note: Box plots for target sample at 25% , 50%, and 100% are not included since the models did not converge or a very few runs successfully fitted the models.

Figure 3.2.2a: Box Plots of the Model Specific Estimates of the Intercepts Obtained from fitting Cross-Sectional Models to 100 Boot Strap Sampling with Target Sampling at 25%, 50%, 100%, 150%, and 250% of the Observed NIS Sample Data for IQI-11.



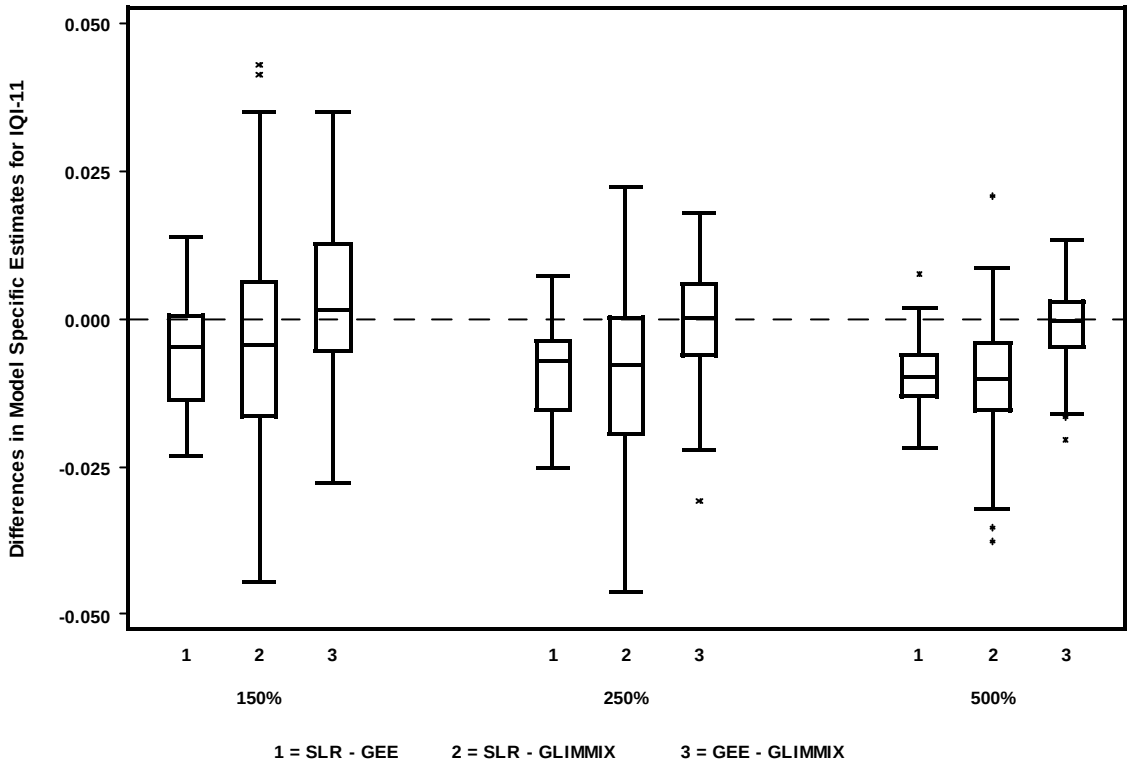
Note: Box plots for target sample at 25% , 50%, and 100% are not included since the models did not converge or a very few runs successfully fitted the models.

Figure 3.2.2b: Box Plots of the Model Specific Estimates of the Intercepts Obtained from fitting Time-Trend Models to 100 Boot Strap Sampling with Target Sampling at 25%, 50%, 100%, 150%, and 250% of the Observed NIS Sample Data for IQI-11.



Note: Box plots for target sample at 25% , 50%, and 100% are not included since the models did not converge or a very few runs successfully fitted the models.

Figure 3.2.2c: Box Plots of the Differences in Model Specific Estimates of the Intercepts from a Specific Pair of Cross-Sectional Models Fitted to 100 Boot Strap Sampling Runs with Target Sampling at 25%, 50%, 100%, 150%, and 250% of the Observed NIS Sample Data for IQI-11.



Note: Box plots for target sample at 25% , 50%, and 100% are not included since the models did not converge or a very few runs successfully fitted the models.

Figure 3.2.2d: Box Plots of the Differences in Model Specific Estimates of the Intercepts from a Specific Pair of Time-Trend Models Fitted to 100 Boot Strap Sampling Runs with Target Sampling at 25%, 50%, 100%, 150%, and 250% of the Observed NIS Sample Data for IQI-11.

Table 3.2.2C Wald Test Statistics and (P-Value) Comparing Models fit to IQI-11 (Abdominal Aortic Artery Repair Mortality Rate)

Test	Percent of Observed Sample	P-values					
		Cross-Sectional Models			Time-Trend Models		
		SLR vs. GEE	SLR vs. GLIM	GEE vs. GLIM	SLR vs. GEE	SLR vs. GLIM	GEE vs. GLIM
Global Wald Test	25%	.	.	.	.	.	.
	50%	.	.	.	.	.	.
	100%	.	.	.	.	.	.
	150%	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
	250%	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
	500%	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
T-Test for Intercept Difference	25%	.	.	.	.	.	.
	50%	.	.	.	.	.	.
	100%	.	.	.	.	.	.
	150%	< 0.0001	< 0.0001	0.4678	< 0.0001	0.3877	0.0275
	250%	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	0.7179
	500%	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	0.0590

3.3 IQI 14 – Hip Replacement Mortality Rate

3.3.1 Model Parameter Estimates from Fitting Models to NIS Sample Data

Below we provide the model parameter estimates from fitting the simple logistic regression, generalized estimating equations, and generalized linear mixed model to the 2001-2003 Nationwide Inpatient Sample for IQI 14 (Hip Replacement Mortality Rate). Table 3.3.1a provides the parameter estimates associated with the cross-sectional model, and Table 3.3.1b provides the parameter estimates associate with the model that adjusts for a simple linear trend over time. Across these two tables, we see that the parameter estimates and associated standard errors are identical among the three modeling approaches, with the estimated correlation coefficient from the GEE modeling approach being estimated as zero in both the cross-sectional and time-trend adjusted models. The estimated variance components associated with provider-specific random effects from the GLIMMIX model was also zero in both models, as well as the variance component that captures provider-specific variation in the time-trend slope was estimated as zero.

Table 3.3.1c provides the Wald Statistics to determine whether there are statistically significant differences between the vector of parameter estimates generated by each modeling approach as applied to the NIS sample. The Wald Statistics consider pair-wise comparisons, and suggest that there were no significant differences between the different modeling approaches for IQI 14.

Table 3.3.1a Parameter Estimates from Cross Sectional Models fit to IQI-14 (Hip Replacement Mortality Rate)

Parameter	Simple Logistic Regression Model			Generalized Estimating Equations Model			Generalized Linear Mixed Model		
	Estimate	Std Err	p value	Estimate	Std Err	p value	Estimate	Std Err	p value
Intercept	-2.377	0.394	0.000	-2.376	0.394	0.000	-2.377	0.394	0.000
SEX	-0.363	0.215	0.091	-0.363	0.215	0.091	-0.363	0.215	0.091
AGE1	-1.573	0.343	0.000	-1.573	0.343	0.000	-1.573	0.343	0.000
AGE10	-1.539	0.404	0.000	-1.539	0.405	0.000	-1.539	0.404	0.000
AGE11	-1.586	0.333	0.000	-1.587	0.333	0.000	-1.586	0.333	0.000
AGE12	-1.017	0.274	0.000	-1.017	0.274	0.000	-1.017	0.274	0.000
AGE13	-0.608	0.252	0.016	-0.608	0.252	0.016	-0.608	0.252	0.016
AGE15	0.549	0.441	0.213	0.549	0.441	0.213	0.549	0.441	0.213
AGE24	0.347	0.543	0.523	0.347	0.543	0.524	0.347	0.543	0.523
AGE25	0.079	0.461	0.865	0.079	0.462	0.864	0.079	0.461	0.865
AGE26	0.142	0.355	0.690	0.141	0.355	0.691	0.142	0.355	0.690
AGE27	-0.092	0.335	0.783	-0.092	0.335	0.783	-0.092	0.335	0.783
C1	-4.016	0.399	0.000	-4.020	0.399	0.000	-4.016	0.399	0.000
C2	-2.260	0.383	0.000	-2.262	0.383	0.000	-2.260	0.383	0.000
C3	0.088	0.378	0.815	0.088	0.378	0.816	0.088	0.378	0.815
C4	2.061	0.384	0.000	2.061	0.384	0.000	2.061	0.384	0.000
C5	0.384	0.559	0.492	0.384	0.559	0.492	0.384	0.559	0.492
ρ				0.0000	.	.			
$\sigma^2_{\text{Btw Hosp}}$							0.000	.	.

The effect of the *YEAR* parameter (which captures the trend over time) was highly significant for all three modeling approaches, as seen in Table 3.3.1b below.

Table 3.3.1b Parameter Estimates from Time Trend Models fit to IQI-14 (Hip Replacement Mortality Rate)

Parameter	Simple Logistic Regression Model			Generalized Estimating Equations Model			Generalized Linear Mixed Model		
	Estimate	Std Err	p value	Estimate	Std Err	p value	Estimate	Std Err	p value
Intercept	-2.385	0.394	0.000	-2.385	0.394	0.000	-2.385	0.394	0.000
SEX	-0.372	0.215	0.083	-0.372	0.215	0.083	-0.372	0.215	0.083
AGE1	-1.570	0.343	0.000	-1.571	0.343	0.000	-1.570	0.343	0.000
AGE10	-1.547	0.404	0.000	-1.547	0.405	0.000	-1.547	0.404	0.000
AGE11	-1.595	0.333	0.000	-1.596	0.334	0.000	-1.595	0.333	0.000
AGE12	-1.027	0.275	0.000	-1.027	0.275	0.000	-1.027	0.275	0.000
AGE13	-0.622	0.252	0.014	-0.622	0.253	0.014	-0.622	0.252	0.014
AGE15	0.564	0.441	0.201	0.564	0.442	0.202	0.564	0.441	0.201
AGE24	0.369	0.543	0.497	0.368	0.544	0.498	0.369	0.543	0.497
AGE25	0.089	0.461	0.846	0.090	0.462	0.845	0.089	0.461	0.846
AGE26	0.156	0.355	0.661	0.155	0.355	0.662	0.156	0.355	0.661
AGE27	-0.070	0.335	0.835	-0.070	0.335	0.834	-0.070	0.335	0.835
C1	-4.013	0.399	0.000	-4.019	0.399	0.000	-4.013	0.399	0.000
C2	-2.250	0.384	0.000	-2.252	0.384	0.000	-2.250	0.384	0.000
C3	0.102	0.378	0.788	0.101	0.378	0.789	0.102	0.378	0.788
C4	2.085	0.384	0.000	2.085	0.384	0.000	2.085	0.384	0.000
C5	0.401	0.560	0.474	0.400	0.560	0.475	0.401	0.560	0.474
YEAR	-0.200	0.076	0.009	-0.201	0.076	0.008	-0.200	0.076	0.009
ρ				-0.0001	.	.			
σ^2_{Hosp}							0.000	.	
σ^2_{Year}							0.000	.	

Table 3.3.1c Wald Test Statistics and (P-Value) Comparing Models fit to IQI-14 (Hip Replacement Mortality Rate)

	Cross Sectional Model			Time Trend Model		
	SLR	GEE	GLIMMIX	SLR	GEE	GLIMMIX
SLR		0.001 (1.000)	0.000 (1.000)		0.002 (1.000)	0.000 (1.000)
GEE	0.001 (1.000)		0.001 (1.000)	0.002 (1.000)		0.002 (1.000)
GLIMMIX	0.000 (1.000)	0.001 (1.000)		0.000 (1.000)	0.002 (1.000)	

- Wald Test uses the estimated covariance matrix from the model listed in each row

Table 3.3.1d below provides the mean and standard deviation of differences between model predictions (expected rates above the diagonal, and risk-adjusted rates below the diagonal) from a random sample of 50 providers within the NIS reference population for IQI 14. Due to the fact that the three modeling approaches yielded identical parameter

estimates (the within-provider correlation was estimated as zero), there were no differences in expected or risk-adjusted rates attributable to the modeling approach.

Table 3.3.1d Estimated Differences (and Standard Deviation) in Provider-Level Model Predictions of Expected and Risk Adjusted Rates for IQI-14 (Hip Replacement Mortality Rate)

	Cross Sectional Model			Time Trend Model		
	SLR	**GEE**	**GLIMMIX**	**SLR**	**GEE**	**GLIMMIX**
SLR		0.000 (0.000)	0.000 (0.000)		0.000 (0.000)	0.000 (0.000)
GEE	-0.000 (0.000)		0.000 (0.000)	-0.000 (0.000)		0.000 (0.000)
GLIMMIX	-0.000 (0.000)	-0.000 (0.000)		-0.000 (0.000)	-0.000 (0.000)	

***** In each 3x3 table above, Expected Rate Differences (and Standard Deviations) are above the diagonal, and Adjusted Rate Differences (and Standard Deviations) are below the diagonal.

3.3.2 Model Parameter Estimates from Fitting Models to Boot Strap Samples of NIS Data

Both cross sectional and time-trend adjusted models (especially GLIMMIX) fitted for quality indicators IQI-19 (Abdominal Aortic Aneurysm Repair Mortality Rate) for bootstrap samples with target sample size of 25% and 50% of the observed NIS sample suffered greatly from convergence problems due to rare occurrence of the response within some of the selected hospitals included in the sample which resulted in convergence problems and degenerate solutions.

Tables 3.3.2a and 3.3.2b provide an overview of the results for the cross sectional and time-trend adjusted models fit to 100 bootstrap samples with population size approximately 100% of the size of the NIS sample. Each table provides the mean across the 100 bootstrap samples of the model parameter estimates and standard errors from the Simple Logistic Regression Model. The table also provides average absolute and relative differences between each pair of modeling approaches. Box plots of model specific intercept estimates and differences in intercept estimates between each pair of modeling approaches when target sample sizes ranging from 50% to 250% of the observed NIS samples for cross-sectional models are displayed in figures 3.3.2a and 3.3.2c. Figures 3.3.2b and 3.3.2d presents the similar boxplots for the time-trend models. The results (p-values) from global Wald tests performed to assess whether there are statistically significant difference between a pair of model parameter estimates are provided in Table 3.3.2c. Results (p-values) from pair-wise T-test to test for intercept differences are also included in Table 3.3.2c. Both Wald and T-tests are performed at 5% level of significance.

Parameter estimates from all three modeling approaches are very close to each other for IQI-14 as seen in Tables 3.3.2a and 3.3.2b. Differences as well as relative differences are very small for all parameters for all pairs of modeling approaches. Variability in parameter estimates are small for target sample size at 250% for all three modeling approaches. The Box plots of intercepts in Figures 3.3.2a and 3.3.2b show comparability among the three modeling approaches, and the Box plots of differences in Figures 3.3.2c and 3.3.2d show that there is less variability and the results are more concentrated around the zero reference line. The Global Wald test for assessing differences among all modeled parameters demonstrates statistically significant differences for all paired modeling approaches with the exception of comparing SLR vs GLIMMIX at 50% target sample. These significant differences are likely attributable to the large sample size and tight variance/covariance estimates among parameter estimates. This drives the inverse of the variance covariance matrix (V^{-1}) towards larger values, which in turn pushes the Wald test statistic to find statistically significant differences among parameter estimates that are seemingly very similar.

Comparisons between intercept estimates for SLR vs. GEE and GEE vs. GLIMMIX using paired t-tests resulted in statistically significant differences for all target sample sizes for time-trend models where as there are no statistically significant differences found for intercept differences between SLR and GLIMMIX which means GEE intercept differences probably are different from SLR and GLIMMIX.

Table 3.3.2a. Average Models Estimates and Relative Standard Errors from Simple Logistic Model Fit, Differences and Relative Differences in Model Estimates from a Specific Pair of Modeling Approaches for the In Patient Quality Indicator IQI-14 (IQI 14 – Hip Replacement Mortality Rate) For Each Parameter in the Model (Cross-Sectional)

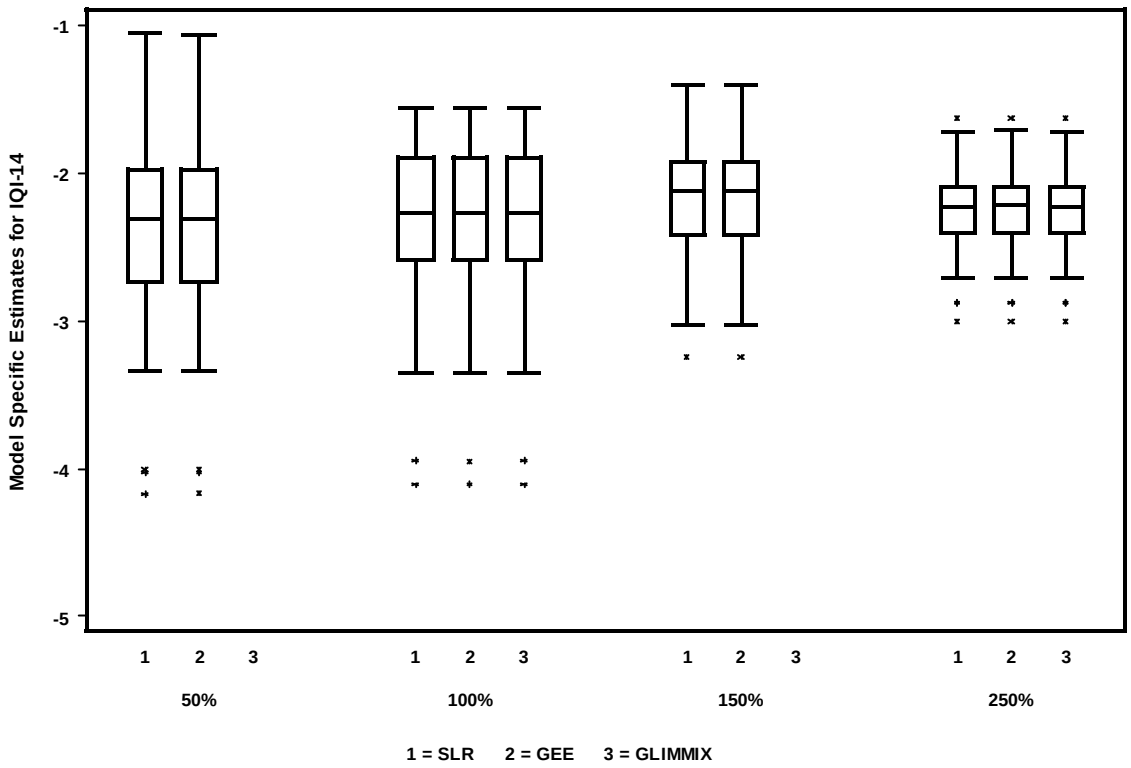
Cross-Sectional								
Parameter	Simple Logistic Regression		Difference between model estimates from a pair of modeling approaches			Relative Difference between model estimates from a pair of modeling approaches		
	Estimate	Std. Error	SLR vs. GEE	SLR vs. GLIM	GEE vs. GLIM	SLR vs. GEE	SLR vs. GLIM	GEE vs. GLIM
Intercept	-2.308	0.4017	-0.001	-0.000	0.000	0.000	0.000	0.000
SEX	-0.370	0.2198	-0.000	-0.001	-0.001	0.000	0.002	0.002
AGE1	-1.487	0.3462	-0.003	-0.000	0.003	0.002	0.000	0.002
AGE10	-1.739	0.4578	-0.005	-0.001	0.004	0.003	0.001	0.002
AGE11	-1.614	0.3479	-0.002	-0.000	0.002	0.001	0.000	0.001
AGE12	-1.079	0.2868	-0.001	-0.001	0.000	0.001	0.001	0.000
AGE13	-0.777	0.2668	-0.001	-0.002	-0.001	0.001	0.002	0.001
AGE15	0.445	0.4550	0.002	0.001	-0.001	0.004	0.003	0.001
AGE24	0.417	0.6213	0.005	0.002	-0.002	0.011	0.005	0.006
AGE25	0.125	0.4806	-0.002	0.001	0.003	0.013	0.010	0.023
AGE26	0.198	0.3685	0.000	0.001	0.001	0.001	0.004	0.004
AGE27	0.028	0.3534	0.001	0.001	0.001	0.024	0.047	0.023
C1	-4.199	0.4108	-0.004	-0.001	0.003	0.001	0.000	0.001
C2	-2.301	0.3903	0.000	-0.001	-0.001	0.000	0.000	0.000
C3	-0.026	0.3862	0.001	-0.000	-0.001	0.030	0.016	0.047
C4	1.918	0.3920	0.001	0.000	-0.001	0.001	0.000	0.001
C5	0.175	0.6196	0.001	0.000	-0.001	0.008	0.002	0.005

Note: Target Sample for the 100 runs is 150% of the Observed NIS sample.

Table 3.3.2b. Average Models Estimates and Relative Standard Errors from Simple Logistic Model Fit, Differences and Relative Differences in Model Estimates from a Specific Pair of Modeling Approaches for the In Patient Quality Indicator IQI-14 (IQI 14 – Hip Replacement Mortality Rate) For Each Parameter in the Model (Cross-Sectional)

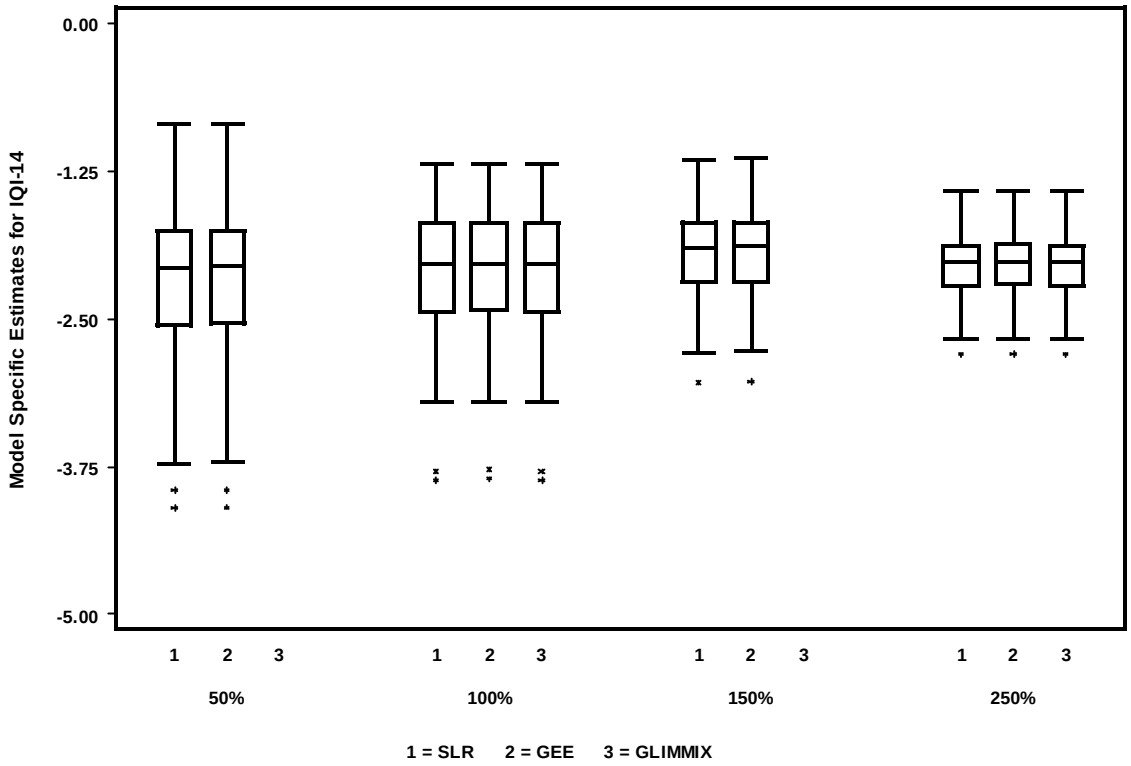
Time -Trend								
Parameter	**Simple Logistic Regression**		**Difference between model estimates from a pair of modeling approaches**			**Relative Difference between model estimates from a pair of modeling approaches**		
	Estimate	**Std. Error**	**SLR vs. GEE**	**SLR vs. GLIM**	**GEE vs. GLIM**	**SLR vs. GEE**	**SLR vs. GLIM**	**GEE vs. GLIM**
Intercept	-2.106	0.4086	-0.007	-0.001	0.006	0.003	0.000	0.003
SEX	-0.380	0.2199	0.000	-0.001	-0.001	0.000	0.002	0.002
AGE1	-1.480	0.3465	-0.003	-0.000	0.003	0.002	0.000	0.002
AGE10	-1.759	0.4583	-0.003	-0.001	0.002	0.002	0.001	0.001
AGE11	-1.625	0.3480	-0.001	-0.001	0.000	0.000	0.000	0.000
AGE12	-1.095	0.2872	-0.000	-0.001	-0.001	0.000	0.001	0.000
AGE13	-0.797	0.2673	-0.000	-0.002	-0.002	0.000	0.003	0.002
AGE15	0.454	0.4552	0.003	0.001	-0.001	0.006	0.003	0.003
AGE24	0.452	0.6217	0.004	0.003	-0.001	0.009	0.006	0.003
AGE25	0.134	0.4809	-0.002	0.002	0.004	0.016	0.011	0.027
AGE26	0.213	0.3689	0.000	0.001	0.001	0.001	0.005	0.004
AGE27	0.058	0.3539	0.000	0.001	0.001	0.001	0.025	0.024
C1	-4.200	0.4110	-0.000	-0.001	-0.000	0.000	0.000	0.000
C2	-2.293	0.3906	0.002	-0.000	-0.002	0.001	0.000	0.001
C3	-0.015	0.3866	0.002	-0.000	-0.002	0.101	0.009	0.111
C4	1.940	0.3925	0.001	0.000	-0.001	0.001	0.000	0.000
C5	0.186	0.6203	0.002	0.001	-0.002	0.013	0.004	0.009
Year	-0.209	0.0782	0.006	0.001	-0.005	0.027	0.003	0.024

Note: Target Sample for the 100 runs is 150% of the Observed NIS sample.



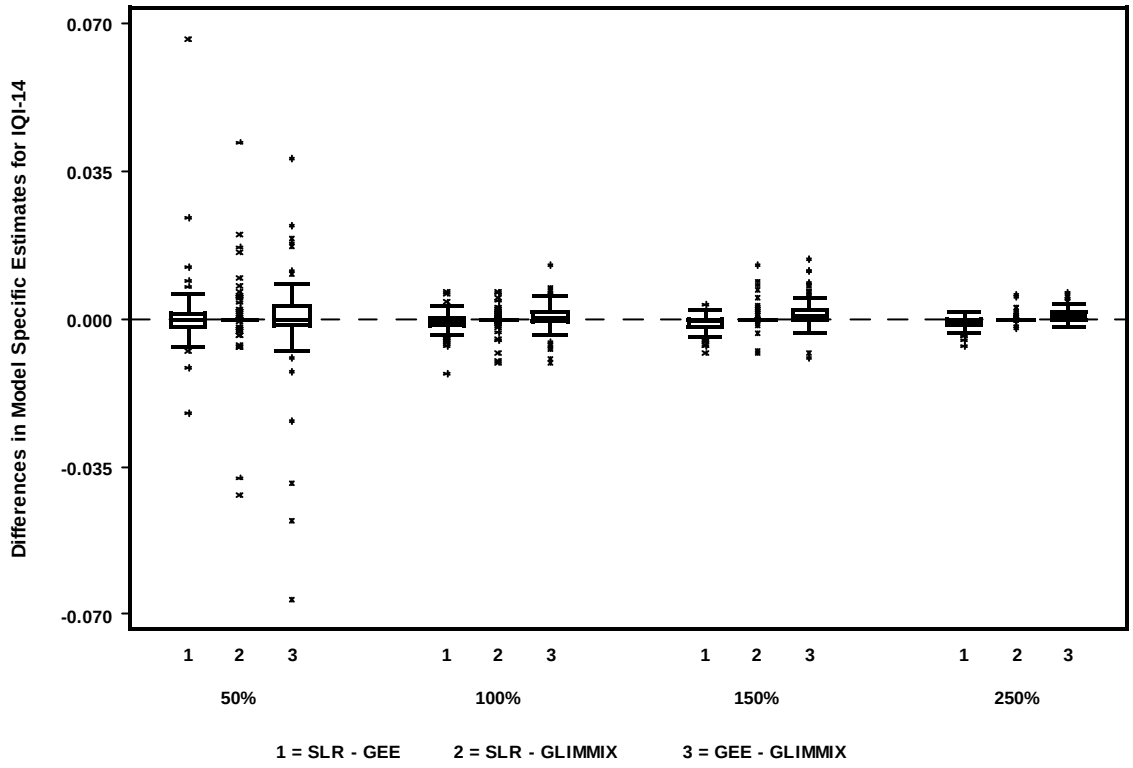
Note: Box plots for target sample at 25% are not included since the models did not converge**.**

Figure 3.3.2a: Box Plots of the Model Specific Estimates of the Intercepts Obtained from fitting Cross-Sectional Models to 100 Boot Strap Sampling with Target Sampling at 25%, 50%, 100%, 150%, and 250% of the Observed NIS Sample Data for IQI-14.



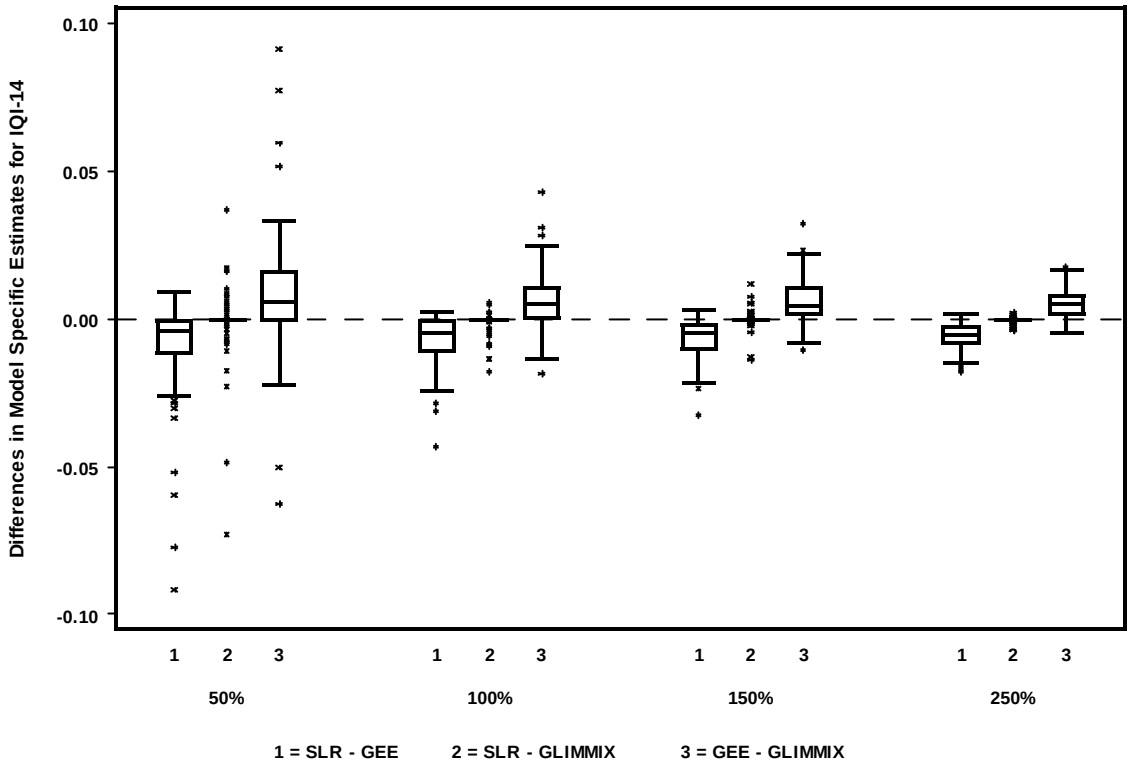
Note: Box plots for target sample at 25% are not included since the models did not converge**.**

Figure 3.3.2b: Box Plots of the Model Specific Estimates of the Intercepts Obtained from fitting Time-Trend Models to 100 Boot Strap Sampling with Target Sampling at 25%, 50%, 100%, 150%, and 250% of the Observed NIS Sample Data for IQI-14.



Note: Box plots for target sample at 25% are not included since the models did not converge**.**

Figure 3.3.2c: Box Plots of the Differences in Model Specific Estimates of the Intercepts from a Specific Pair of Cross-Sectional Models Fitted to 100 Boot Strap Sampling Runs with Target Sampling at 25%, 50%, 100%, 150%, and 250% of the Observed NIS Sample Data for IQI-14.



Note: Box plots for target sample at 25% are not included since the models did not converge**.**

Figure 3.3.2d: Box Plots of the Differences in Model Specific Estimates of the Intercepts from a Specific Pair of Time-Trend Models Fitted to 100 Boot Strap Sampling Runs with Target Sampling at 25%, 50%, 100%, 150%, and 250% of the Observed NIS Sample Data for IQI-14.

Table 3.3.2C. Wald Test Statistics and (P-Value) Comparing Models fit to IQI-14 (IQI 14 – Hip Replacement Mortality Rate)

Test	Percent of Observed Sample	P-values					
		Cross-Sectional Models			Time-Trend Models		
		SLR vs. GEE	SLR vs. GLIM	GEE vs. GLIM	SLR vs. GEE	SLR vs. GLIM	GEE vs. GLIM
Global Wald Test	25%	.	.	.	.	.	.
	50%	< 0.0001	0.4284	< 0.0001	< 0.0001	< 0.0001	0.0002
	100%	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	0.02091
	150%	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
	250%	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
T-Test for Intercept Difference	25%	.	.	.	.	.	.
	50%	0.5098	0.5992	0.9256	< 0.0001	0.4893	0.0001
	100%	0.0252	0.7000	0.1580	< 0.0001	0.0144	< 0.0001
	150%	< 0.0001	0.0931	< 0.0001	< 0.0001	0.9966	< 0.0001
	250%	< 0.0001	0.0347	< 0.0001	< 0.0001	0.0499	< 0.0001

3.4 IQI 17 – Acute Stroke Mortality Rate

3.4.1 Model Parameter Estimates from Fitting Models to NIS Sample Data

Below we provide the model parameter estimates from fitting the simple logistic regression, generalized estimating equations, and generalized linear mixed model to the 2001-2003 Nationwide Inpatient Sample for IQI 17 (Acute Stroke Mortality Rate). Table 3.4.1a provides the parameter estimates associated with the cross-sectional model, and Table 3.4.1b provides the parameter estimates associate with the model that adjusts for a simple linear trend over time. Across these two tables, we see that the parameter estimates and associated standard errors are comparable among the three modeling approaches (but with noticeable differences). In this model, the estimated correlation coefficient from the GEE modeling approach is nearly zero in both the cross-sectional (ρ=0.0082) and time-trend adjusted (ρ=0.0079) models. The estimated variance components associated with provider-specific random effects from the GLIMMIX model were small, yet statistically significant in both models ($\sigma^2_{\text{Btw Hosp}} = 0.173$ for the cross sectional model, and 0.161 for the time-trend model). The variance component that captures provider-specific variation in the time-trend slope was estimated as statistically significant ($\sigma^2_{\text{Year}} = 0.019$).

The effect of the *YEAR* parameter (which captures the trend over time) was highly significant for all three modeling approaches, as seen in Table 3.4.1b below.

Table 3.4.1c provides the Wald Statistics to determine whether there are statistically significant differences between the vector of parameter estimates generated by each modeling approach. The Wald Statistics consider pair-wise comparisons, and suggest that there were significant differences between the different modeling approaches for IQI 17. In fact, the first row of Table 3.2.1c suggests that there are highly significant differences between the parameter estimates from the simple logistic regression model compared to both approaches (GEE and GLIMMIX) that adjust for the potential within-hospital correlation among responses. This first row corresponds to a Wald Test that uses the estimated variance/covariance matrix from the simple logistic regression model. The corresponding Wald-Statistic Results that utilize the variance-covariance matrices from the GEE or GLIMMIX approaches are contained in the SLR columns (on each side of the table) – and interestingly enough do not meet the threshold of being significantly different. These variance/covariance matrices are model-based and may be subject to model misspecification. A subsequent iteration of this report will integrate similar Wald Test statistics using the robust variance/covariance estimates for the GEE and GLIMMIX approaches.

Another interesting phenomenon is that there are statistically significant differences between the GEE and GLIMMIX approaches (for both the cross-sectional and time-trend models). Again, this result necessitates further diagnoses of the fit of these models.

Table 3.4.1a Parameter Estimates from Cross Sectional Models fit to IQI-17 (Acute Stroke Mortality Rate)

Parameter	Simple Logistic Regression Model			Generalized Estimating Equations Model			Generalized Linear Mixed Model		
	Estimate	Std Err	p value	Estimate	Std Err	p value	Estimate	Std Err	p value
Intercept	-1.727	0.068	0.000	-1.655	0.066	0.000	-1.660	0.069	0.000
SEX	0.067	0.028	0.018	0.063	0.028	0.024	0.063	0.028	0.028
AGE1	-0.951	0.177	0.000	-0.910	0.173	0.000	-0.945	0.180	0.000
AGE2	-0.994	0.168	0.000	-0.935	0.163	0.000	-0.952	0.170	0.000
AGE3	-0.682	0.110	0.000	-0.640	0.107	0.000	-0.663	0.112	0.000
AGE4	-0.663	0.090	0.000	-0.619	0.088	0.000	-0.636	0.091	0.000
AGE5	-0.669	0.068	0.000	-0.636	0.067	0.000	-0.662	0.069	0.000
AGE6	-0.472	0.055	0.000	-0.439	0.054	0.000	-0.454	0.056	0.000
AGE7	-0.483	0.049	0.000	-0.454	0.048	0.000	-0.473	0.049	0.000
AGE8	-0.453	0.047	0.000	-0.421	0.046	0.000	-0.437	0.048	0.000
AGE9	-0.442	0.044	0.000	-0.411	0.043	0.000	-0.422	0.045	0.000
AGE10	-0.375	0.040	0.000	-0.352	0.039	0.000	-0.362	0.041	0.000
AGE11	-0.419	0.037	0.000	-0.393	0.036	0.000	-0.405	0.037	0.000
AGE12	-0.329	0.034	0.000	-0.311	0.033	0.000	-0.320	0.035	0.000
AGE13	-0.208	0.034	0.000	-0.197	0.033	0.000	-0.205	0.034	0.000
AGE15	-0.366	0.279	0.190	-0.336	0.271	0.215	-0.336	0.284	0.236
AGE16	-0.211	0.252	0.403	-0.203	0.244	0.405	-0.235	0.255	0.356
AGE17	-0.304	0.162	0.060	-0.287	0.157	0.068	-0.292	0.163	0.075
AGE18	-0.252	0.129	0.051	-0.243	0.126	0.053	-0.258	0.131	0.048
AGE19	0.126	0.093	0.177	0.127	0.091	0.163	0.134	0.094	0.155
AGE20	-0.114	0.077	0.140	-0.107	0.075	0.155	-0.110	0.078	0.160
AGE21	-0.149	0.069	0.030	-0.131	0.067	0.049	-0.125	0.069	0.073
AGE22	-0.150	0.066	0.023	-0.149	0.064	0.021	-0.152	0.067	0.022
AGE23	-0.049	0.062	0.425	-0.050	0.060	0.408	-0.051	0.062	0.411
AGE24	-0.138	0.055	0.013	-0.134	0.054	0.013	-0.139	0.056	0.013
AGE25	-0.014	0.048	0.764	-0.018	0.047	0.700	-0.018	0.049	0.716
AGE26	-0.045	0.044	0.301	-0.041	0.043	0.342	-0.039	0.044	0.380
AGE27	-0.087	0.042	0.039	-0.083	0.041	0.044	-0.083	0.042	0.050
C1	-1.548	0.140	0.000	-1.465	0.130	0.000	-1.558	0.141	0.000
C2	-0.111	0.082	0.175	-0.113	0.079	0.154	-0.114	0.083	0.169
C3	1.019	0.070	0.000	0.986	0.068	0.000	1.022	0.071	0.000
C4	2.289	0.074	0.000	2.260	0.072	0.000	2.341	0.075	0.000
C5	-0.211	0.085	0.013	-0.222	0.082	0.007	-0.236	0.086	0.006
C6	0.411	0.066	0.000	0.377	0.064	0.000	0.383	0.067	0.000
C7	1.578	0.068	0.000	1.535	0.066	0.000	1.580	0.069	0.000
C8	3.312	0.068	0.000	3.290	0.066	0.000	3.375	0.069	0.000
C9	-2.565	0.077	0.000	-2.425	0.074	0.000	-2.610	0.077	0.000
C10	-1.224	0.066	0.000	-1.204	0.064	0.000	-1.264	0.067	0.000
C11	0.295	0.067	0.000	0.277	0.064	0.000	0.284	0.067	0.000
C12	2.126	0.067	0.000	2.089	0.065	0.000	2.152	0.067	0.000
C13	-2.451	0.094	0.000	-2.387	0.090	0.000	-2.525	0.095	0.000
C14	-0.958	0.068	0.000	-0.984	0.066	0.000	-1.030	0.069	0.000
C15	0.359	0.072	0.000	0.315	0.070	0.000	0.315	0.073	0.000
C16	2.031	0.078	0.000	1.987	0.076	0.000	2.047	0.079	0.000
C17	-0.014	0.072	0.843	-0.010	0.069	0.880	-0.007	0.073	0.922
ρ				0.0082	.	.			
σ^2_{Hosp}							0.173	0.011	.

Table 3.4.1b Parameter Estimates from Time Trend Models fit to IQI-17 (Acute Stroke Mortality Rate)

Parameter	Simple Logistic Regression Model			Generalized Estimating Equations Model			Generalized Linear Mixed Model		
	Estimate	Std Err	p value	Estimate	Std Err	p value	Estimate	Std Err	p value
Intercept	-1.725	0.068	0.000	-1.656	0.066	0.000	-1.662	0.069	0.000
SEX	0.066	0.028	0.020	0.062	0.028	0.025	0.062	0.028	0.029
AGE1	-0.951	0.177	0.000	-0.913	0.173	0.000	-0.944	0.180	0.000
AGE2	-0.990	0.168	0.000	-0.936	0.163	0.000	-0.951	0.170	0.000
AGE3	-0.681	0.110	0.000	-0.640	0.107	0.000	-0.664	0.112	0.000
AGE4	-0.661	0.090	0.000	-0.620	0.088	0.000	-0.638	0.091	0.000
AGE5	-0.669	0.068	0.000	-0.637	0.067	0.000	-0.663	0.069	0.000
AGE6	-0.468	0.055	0.000	-0.438	0.054	0.000	-0.454	0.056	0.000
AGE7	-0.480	0.049	0.000	-0.453	0.048	0.000	-0.474	0.049	0.000
AGE8	-0.451	0.047	0.000	-0.421	0.046	0.000	-0.438	0.048	0.000
AGE9	-0.441	0.045	0.000	-0.412	0.043	0.000	-0.423	0.045	0.000
AGE10	-0.375	0.040	0.000	-0.353	0.039	0.000	-0.363	0.041	0.000
AGE11	-0.420	0.037	0.000	-0.395	0.036	0.000	-0.405	0.037	0.000
AGE12	-0.331	0.034	0.000	-0.313	0.033	0.000	-0.322	0.035	0.000
AGE13	-0.209	0.034	0.000	-0.198	0.033	0.000	-0.205	0.034	0.000
AGE15	-0.352	0.279	0.207	-0.326	0.271	0.229	-0.334	0.284	0.240
AGE16	-0.207	0.252	0.411	-0.201	0.244	0.410	-0.232	0.255	0.363
AGE17	-0.307	0.162	0.058	-0.290	0.158	0.066	-0.291	0.164	0.075
AGE18	-0.257	0.129	0.046	-0.246	0.126	0.050	-0.258	0.131	0.048
AGE19	0.129	0.093	0.167	0.129	0.091	0.157	0.135	0.094	0.154
AGE20	-0.113	0.077	0.142	-0.107	0.075	0.156	-0.110	0.078	0.160
AGE21	-0.150	0.069	0.029	-0.132	0.067	0.049	-0.123	0.069	0.076
AGE22	-0.150	0.066	0.023	-0.149	0.064	0.021	-0.152	0.067	0.023
AGE23	-0.048	0.062	0.433	-0.049	0.060	0.415	-0.050	0.062	0.424
AGE24	-0.137	0.055	0.013	-0.133	0.054	0.014	-0.139	0.056	0.013
AGE25	-0.014	0.048	0.770	-0.018	0.047	0.705	-0.018	0.049	0.708
AGE26	-0.044	0.044	0.317	-0.040	0.043	0.354	-0.038	0.044	0.386
AGE27	-0.086	0.042	0.039	-0.083	0.041	0.044	-0.083	0.042	0.049
C1	-1.550	0.140	0.000	-1.468	0.130	0.000	-1.558	0.141	0.000
C2	-0.112	0.082	0.173	-0.113	0.079	0.155	-0.113	0.083	0.174
C3	1.016	0.070	0.000	0.986	0.068	0.000	1.021	0.071	0.000
C4	2.291	0.074	0.000	2.264	0.072	0.000	2.344	0.075	0.000
C5	-0.210	0.085	0.014	-0.221	0.082	0.007	-0.234	0.086	0.006
C6	0.409	0.066	0.000	0.378	0.064	0.000	0.383	0.067	0.000
C7	1.579	0.068	0.000	1.538	0.066	0.000	1.582	0.069	0.000
C8	3.314	0.068	0.000	3.293	0.066	0.000	3.378	0.069	0.000
C9	-2.568	0.077	0.000	-2.429	0.074	0.000	-2.609	0.077	0.000
C10	-1.225	0.066	0.000	-1.204	0.064	0.000	-1.263	0.067	0.000
C11	0.297	0.067	0.000	0.280	0.065	0.000	0.286	0.067	0.000
C12	2.128	0.067	0.000	2.093	0.065	0.000	2.155	0.067	0.000
C13	-2.458	0.094	0.000	-2.391	0.090	0.000	-2.526	0.095	0.000
C14	-0.965	0.068	0.000	-0.987	0.066	0.000	-1.030	0.069	0.000
C15	0.353	0.072	0.000	0.313	0.070	0.000	0.314	0.073	0.000
C16	2.026	0.078	0.000	1.986	0.076	0.000	2.047	0.079	0.000
C17	-0.011	0.072	0.874	-0.009	0.070	0.901	-0.005	0.073	0.945
YEAR	-0.065	0.008	0.000	-0.053	0.010	0.000	-0.059	0.013	0.000
ρ				0.0079	.	.			
σ^2_{Hosp}							0.161	0.011	.
σ^2_{Year}							0.019	0.007	.

Table 3.4.1c Wald Test Statistics and (P-Value) Comparing Models fit to IQI-17 (Acute Stroke Mortality Rate)

	Cross Sectional Model			Time Trend Model		
	SLR	**GEE**	**GLIMMIX**	**SLR**	**GEE**	**GLIMMIX**
SLR		154.85 (0.000)	127.77 (0.000)		149.27 (0.000)	120.62 (0.000)
GEE	57.362 (0.102)		64.972 (0.027)	56.246 (0.143)		62.365 (0.054)
GLIMMIX	49.287 (0.306)	70.869 (0.008)		46.309 (0.460)	68.044 (0.019)	

***** Wald Test uses the estimated covariance matrix from the model listed in each row

Table 3.4.1d below provides the mean and standard deviation of differences between model predictions (expected rates above the diagonal, and risk-adjusted rates below the diagonal) from a random sample of 50 providers within the NIS reference population for IQI 17. For example, the mean difference in expected rates between the simple logistic regression and GEE approaches for the cross sectional model was -0.006 (relative to a national mean response rate of 0.114 from Table 3.1). Mean differences (and standard deviations) attributable to model specification (simple logistic vs GEE vs GLIMMIX) for the risk-adjusted rates appear to be higher than the expected rates.

Table 3.4.1d Estimated Differences (and Standard Deviation) in Provider-Level Model Predictions of Expected and Risk Adjusted Rates for IQI-17 (Acute Stroke Mortality Rate)

	Cross Sectional Model			Time Trend Model		
	SLR	**GEE**	**GLIMMIX**	**SLR**	**GEE**	**GLIMMIX**
SLR		-0.006 (0.001)	-0.002 (0.001)		-0.005 (0.001)	-0.002 (0.001)
GEE	0.021 (0.015)		0.004 (0.000)	0.021 (0.015)		0.004 (0.001)
GLIMMIX	0.007 (0.004)	-0.014 (0.011)		0.007 (0.005)	-0.014 (0.011)	

***** In each 3x3 table above, Expected Rate Differences (and Standard Deviations) are above the diagonal, and Adjusted Rate Differences (and Standard Deviations) are below the diagonal.

3.4.2 Model Parameter Estimates from Fitting Models to Boot Strap Samples of NIS Data

Both cross sectional and time-trend adjusted models (especially GLIMMIX) fitted for quality indicators IQI-19 (Abdominal Aortic Aneurysm Repair Mortality Rate) for bootstrap sampling runs with target sample size of 25% and 50% of the observed NIS sample had convergence problems due to small sample size or rare occurrence of the event.

Tables 3.4.2a and 3.4.2b provide an overview of the results for the cross sectional and time-trend adjusted models fit to 100 bootstrap samples with population size approximately 100% of the size of the NIS sample. Each table provides the mean across the 100 bootstrap samples of the model parameter estimates and standard errors from the Simple Logistic Regression Model. The table also provides average absolute and relative differences between each pair of modeling approaches. Box plots of model specific intercept estimates and differences in intercept estimates between each pair of modeling approaches when target sample sizes ranging from 25% to 250% of the observed NIS samples for cross-sectional models are displayed in figures 3.4.2a and 3.4.2c. Figures 3.4.2b and 3.4.2d presents the similar boxplots for the time-trend models. The results (p-values) from global Wald tests performed to assess whether there are statistically significant difference between a pair of model parameter estimates are provided in Table 3.4.2c. Results (p-values) from pair-wise T-test to test for intercept differences are also included in Table 3.4.2c. Both Wald and T-tests are performed at 5% level of significance.

All three model specific parameter estimates are quite comparable and the variability decreases as target sample size increases. For both cross-sectional and time trend models the intercept differences between GEE and GLIMMIX models are always the highest for all target sample sizes as seen from figures 3.4.2c and 3.4.2d. All global Wald tests and t-tests for intercept differences are highly significant which means the three modeling approaches are statistically different.

Table 3.4.2a. Average Models Estimates and Relative Standard Errors from Simple Logistic Model Fit, Differences and Relative Differences in Model Estimates from a Specific Pair of Modeling Approaches for the In Patient Quality Indicator IQI-17 (IQI 17 – Acute Stroke Mortality Rate) For Each Parameter in the Model (Cross-Sectional)

Cross-Sectional								
Parameter	Simple Logistic Regression		Difference between model estimates from a pair of modeling approaches			Relative Difference between model estimates from a pair of modeling approaches		
	Estimate	Std. Error	SLR vs. GEE	SLR vs. GLIM	GEE vs. GLIM	SLR vs. GEE	SLR vs. GLIM	GEE vs. GLIM
Intercept	-1.727	0.0671	-0.053	-0.037	0.016	0.031	0.022	0.009
SEX	0.077	0.0286	0.003	0.002	-0.001	0.033	0.022	0.011
AGE1	-0.896	0.1742	-0.033	-0.004	0.030	0.038	0.004	0.034
AGE2	-0.961	0.1700	-0.053	-0.040	0.013	0.057	0.043	0.015
AGE3	-0.634	0.1097	-0.037	-0.018	0.019	0.060	0.028	0.032
AGE4	-0.635	0.0890	-0.036	-0.022	0.014	0.059	0.035	0.024
AGE5	-0.669	0.0676	-0.027	-0.001	0.026	0.042	0.002	0.040
AGE6	-0.440	0.0548	-0.026	-0.011	0.015	0.061	0.026	0.036
AGE7	-0.475	0.0485	-0.023	-0.004	0.019	0.050	0.008	0.042
AGE8	-0.416	0.0467	-0.027	-0.013	0.014	0.068	0.031	0.036
AGE9	-0.416	0.0443	-0.027	-0.018	0.009	0.067	0.044	0.023
AGE10	-0.352	0.0403	-0.021	-0.012	0.008	0.061	0.036	0.025
AGE11	-0.407	0.0367	-0.021	-0.011	0.010	0.054	0.027	0.027
AGE12	-0.314	0.0344	-0.015	-0.008	0.007	0.049	0.025	0.025
AGE13	-0.195	0.0339	-0.009	-0.003	0.006	0.049	0.018	0.031
AGE15	-0.404	0.2783	-0.033	-0.034	-0.001	0.084	0.088	0.004
AGE16	-0.324	0.2563	-0.004	0.032	0.036	0.012	0.095	0.107
AGE17	-0.391	0.1621	-0.014	-0.006	0.008	0.036	0.016	0.020
AGE18	-0.279	0.1274	-0.008	0.007	0.015	0.029	0.024	0.053
AGE19	0.133	0.0925	-0.003	-0.011	-0.009	0.021	0.082	0.062
AGE20	-0.143	0.0765	-0.007	-0.004	0.003	0.049	0.027	0.023
AGE21	-0.154	0.0684	-0.016	-0.024	-0.007	0.111	0.166	0.055
AGE22	-0.181	0.0655	0.000	0.005	0.005	0.001	0.026	0.025
AGE23	-0.085	0.0616	0.002	0.007	0.005	0.027	0.079	0.052
AGE24	-0.157	0.0554	-0.001	0.005	0.007	0.008	0.034	0.042
AGE25	-0.012	0.0484	0.004	0.005	0.001	0.320	0.358	0.039
AGE26	-0.060	0.0440	-0.003	-0.003	-0.000	0.046	0.053	0.007
AGE27	-0.099	0.0424	-0.003	-0.002	0.001	0.026	0.017	0.009
C1	-1.506	0.1349	-0.052	0.013	0.065	0.035	0.009	0.044
C2	-0.128	0.0810	0.004	0.006	0.002	0.029	0.046	0.017
C3	0.987	0.0687	0.026	-0.000	-0.026	0.026	0.000	0.027
C4	2.268	0.0727	0.020	-0.049	-0.069	0.009	0.021	0.030
C5	-0.226	0.0843	0.011	0.023	0.012	0.047	0.095	0.049
C6	0.375	0.0651	0.027	0.023	-0.004	0.075	0.063	0.012
C7	1.568	0.0674	0.033	-0.005	-0.038	0.021	0.003	0.025
C8	3.299	0.0672	0.015	-0.062	-0.077	0.005	0.019	0.023
C9	-2.601	0.0757	-0.100	0.038	0.138	0.039	0.015	0.054
C10	-1.244	0.0652	-0.013	0.034	0.047	0.010	0.027	0.037
C11	0.271	0.0659	0.012	0.006	-0.006	0.044	0.021	0.023
C12	2.118	0.0659	0.027	-0.027	-0.054	0.013	0.013	0.026
C13	-2.574	0.0971	-0.038	0.062	0.101	0.015	0.024	0.039
C14	-1.017	0.0677	0.026	0.060	0.034	0.025	0.057	0.032
C15	0.326	0.0717	0.035	0.034	-0.001	0.114	0.110	0.004
C16	2.039	0.0772	0.033	-0.021	-0.054	0.016	0.010	0.027
C17	-0.024	0.0709	-0.003	-0.005	-0.002	0.118	0.235	0.118

Table 3.4.2 b. Average Models Estimates and Relative Standard Errors from Simple Logistic Model Fit, Differences and Relative Differences in Model Estimates from a Specific Pair of Modeling Approaches for the In Patient Quality Indicator IQI-17 (IQI 17 – Acute Stroke Mortality Rate) For Each Parameter in the Model (Time-Trend)

Parameter	Simple Logistic Regression		Difference between model estimates from a pair of modeling approaches			Relative Difference between model estimates from a pair of modeling approaches		
	Estimate	Std. Error	SLR vs. GEE	SLR vs. GLIM	GEE vs. GLIM	SLR vs. GEE	SLR vs. GLIM	GEE vs. GLIM
Intercept	-1.657	0.0676	-0.035	-0.019	0.016	0.021	0.012	0.010
SEX	0.075	0.0286	0.002	0.001	-0.001	0.028	0.015	0.013
AGE1	-0.896	0.1744	-0.030	-0.001	0.029	0.034	0.001	0.034
AGE2	-0.957	0.1700	-0.049	-0.035	0.014	0.051	0.036	0.015
AGE3	-0.632	0.1096	-0.035	-0.016	0.018	0.055	0.026	0.031
AGE4	-0.633	0.0891	-0.033	-0.020	0.014	0.053	0.031	0.023
AGE5	-0.668	0.0676	-0.026	-0.001	0.025	0.038	0.001	0.039
AGE6	-0.435	0.0548	-0.023	-0.008	0.015	0.053	0.018	0.037
AGE7	-0.472	0.0485	-0.020	-0.001	0.019	0.043	0.002	0.042
AGE8	-0.414	0.0467	-0.025	-0.010	0.014	0.060	0.025	0.037
AGE9	-0.414	0.0443	-0.025	-0.016	0.009	0.060	0.038	0.023
AGE10	-0.352	0.0403	-0.019	-0.011	0.008	0.055	0.032	0.025
AGE11	-0.409	0.0368	-0.021	-0.011	0.010	0.052	0.028	0.026
AGE12	-0.316	0.0344	-0.016	-0.009	0.007	0.050	0.028	0.023
AGE13	-0.196	0.0339	-0.009	-0.004	0.006	0.047	0.018	0.030
AGE15	-0.389	0.2784	-0.027	-0.028	-0.001	0.070	0.071	0.002
AGE16	-0.319	0.2564	-0.001	0.035	0.036	0.004	0.109	0.113
AGE17	-0.393	0.1622	-0.014	-0.007	0.007	0.036	0.019	0.018
AGE18	-0.285	0.1274	-0.010	0.003	0.014	0.036	0.012	0.050
AGE19	0.137	0.0926	-0.001	-0.008	-0.007	0.008	0.061	0.053
AGE20	-0.143	0.0765	-0.006	-0.003	0.003	0.044	0.022	0.024
AGE21	-0.155	0.0684	-0.017	-0.024	-0.007	0.108	0.155	0.053
AGE22	-0.181	0.0655	0.000	0.005	0.004	0.001	0.025	0.024
AGE23	-0.084	0.0616	0.003	0.007	0.005	0.031	0.088	0.055
AGE24	-0.157	0.0554	-0.001	0.005	0.007	0.008	0.035	0.043
AGE25	-0.011	0.0484	0.005	0.006	0.001	0.424	0.511	0.061
AGE26	-0.058	0.0440	-0.002	-0.002	-0.000	0.031	0.032	0.000
AGE27	-0.099	0.0424	-0.002	-0.001	0.001	0.024	0.014	0.009
C1	-1.508	0.1349	-0.051	0.011	0.063	0.034	0.007	0.043
C2	-0.128	0.0810	0.003	0.005	0.002	0.024	0.039	0.015
C3	0.985	0.0687	0.024	-0.002	-0.025	0.024	0.002	0.026
C4	2.271	0.0728	0.019	-0.048	-0.067	0.008	0.021	0.030
C5	-0.225	0.0843	0.011	0.023	0.012	0.048	0.100	0.050
C6	0.373	0.0651	0.025	0.021	-0.004	0.067	0.056	0.012
C7	1.569	0.0674	0.032	-0.006	-0.037	0.020	0.004	0.024
C8	3.301	0.0672	0.014	-0.061	-0.076	0.004	0.019	0.023
C9	-2.605	0.0757	-0.099	0.035	0.133	0.038	0.013	0.053
C10	-1.246	0.0652	-0.014	0.031	0.045	0.011	0.025	0.037
C11	0.272	0.0659	0.011	0.005	-0.006	0.040	0.018	0.023
C12	2.120	0.0659	0.026	-0.027	-0.053	0.012	0.013	0.025
C13	-2.582	0.0971	-0.042	0.056	0.098	0.016	0.022	0.039
C14	-1.025	0.0677	0.020	0.053	0.033	0.020	0.052	0.032
C15	0.319	0.0717	0.031	0.029	-0.002	0.096	0.090	0.007
C16	2.034	0.0773	0.029	-0.025	-0.054	0.014	0.012	0.027
C17	-0.021	0.0709	-0.002	-0.004	-0.002	0.076	0.179	0.111
Year	-0.068	0.0080	-0.015	-0.015	-0.000	0.217	0.218	0.002

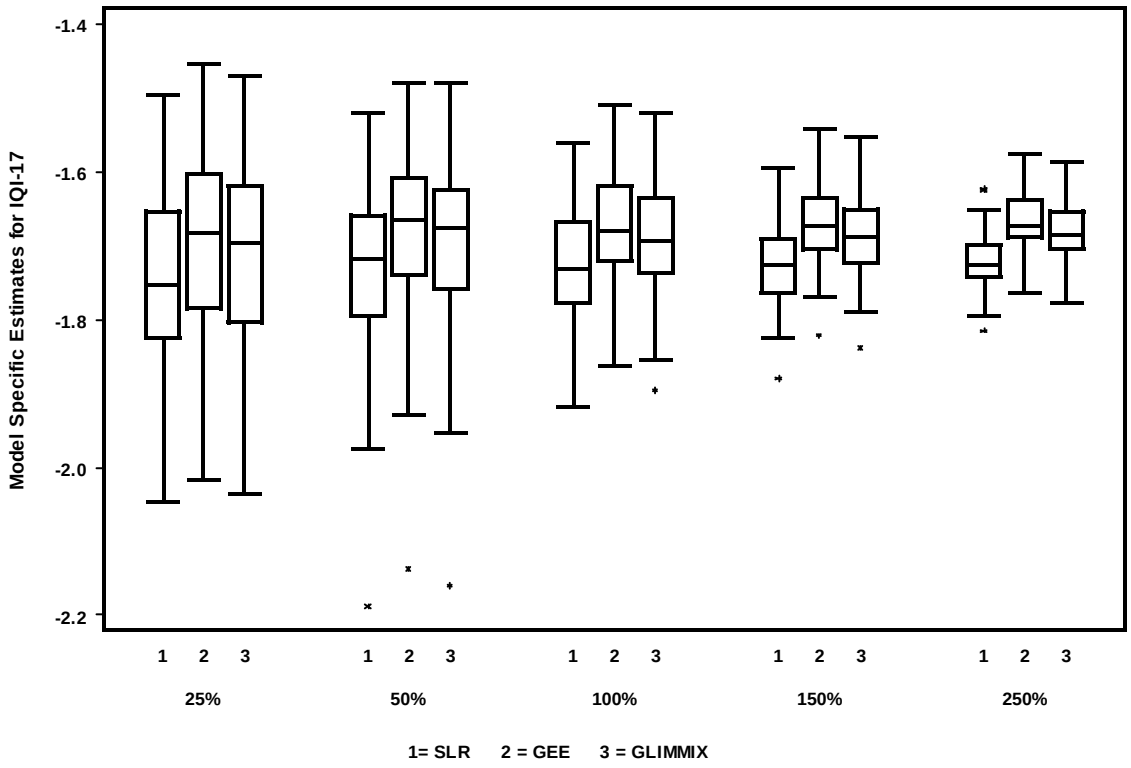


Figure 3.4.2a: Box Plots of the Model Specific Estimates of the Intercepts Obtained from fitting Cross-Sectional Models to 100 Boot Strap Sampling with Target Sampling at 25%, 50%, 100%, 150%, and 250% of the Observed NIS Sample Data for IQI-17.

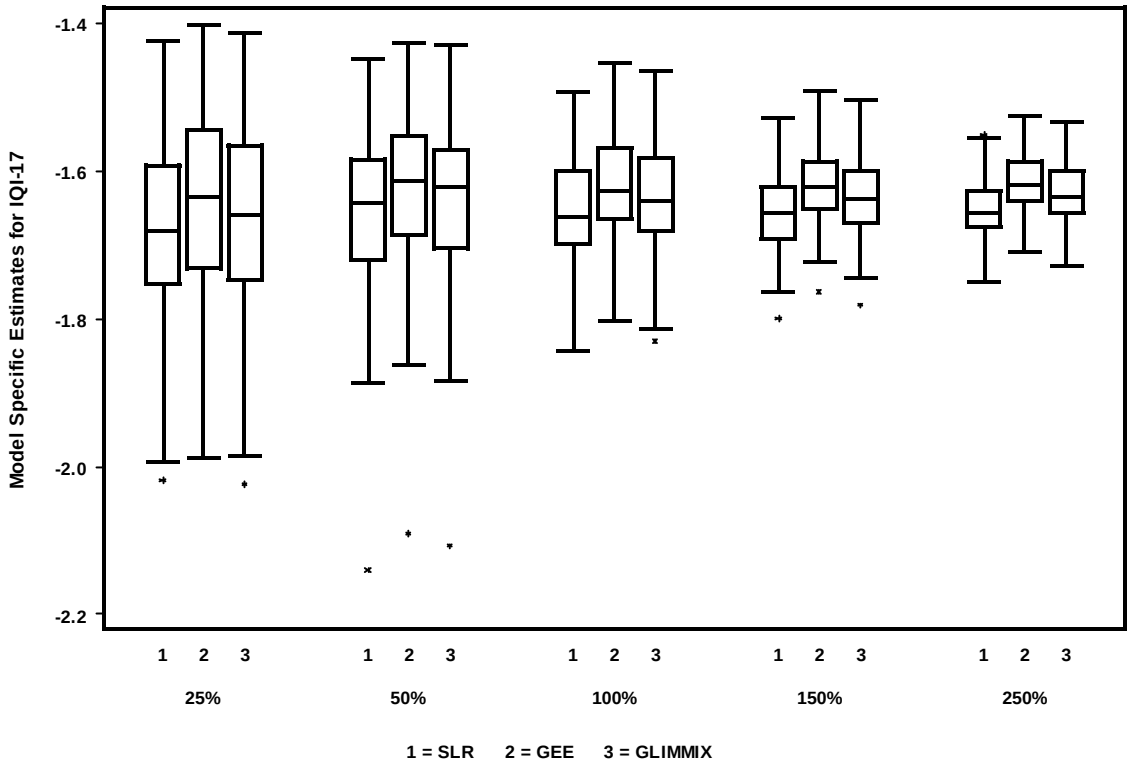


Figure 3.4.2b: Box Plots of the Model Specific Estimates of the Intercepts Obtained from fitting Time-Trend Models to 100 Boot Strap Sampling with Target Sampling at 25%, 50%, 100%, 150%, and 250% of the Observed NIS Sample Data for IQI-17.

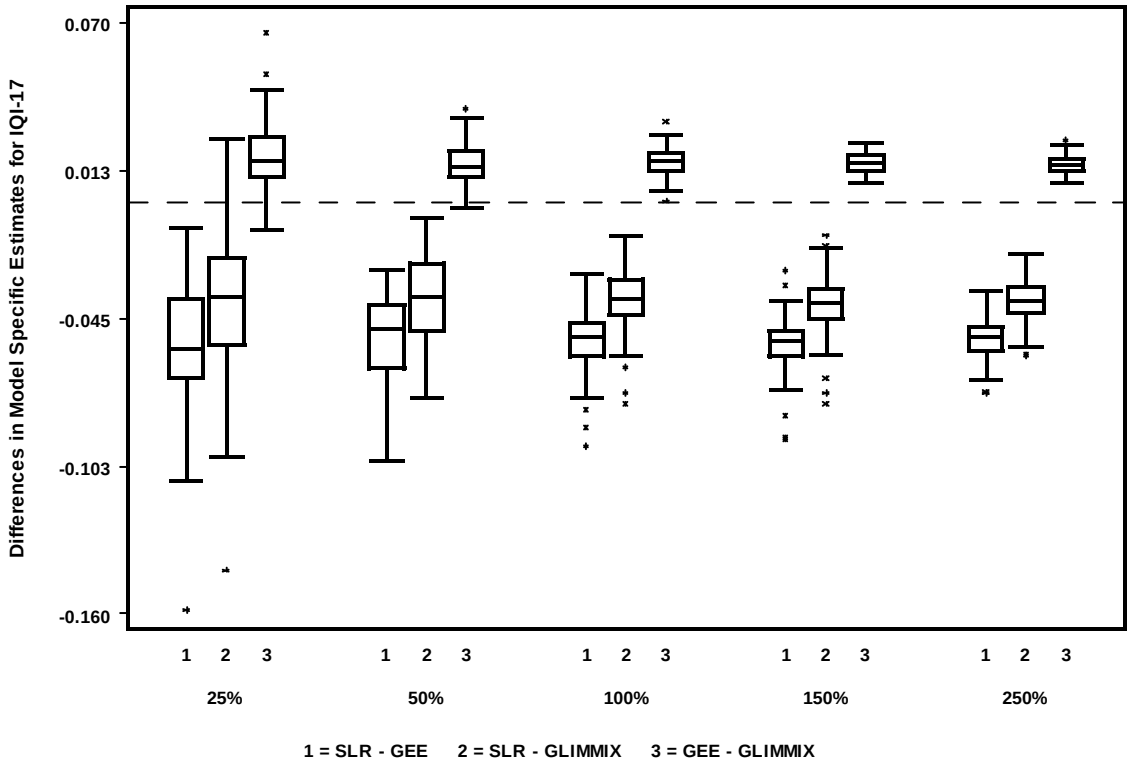


Figure 3.4.2c: Box Plots of the Differences in Model Specific Estimates of the Intercepts from a Specific Pair of Cross-Sectional Models Fitted to 100 Boot Strap Sampling Runs with Target Sampling at 25%, 50%, 100%, 150%, and 250% of the Observed NIS Sample Data for IQI-17.

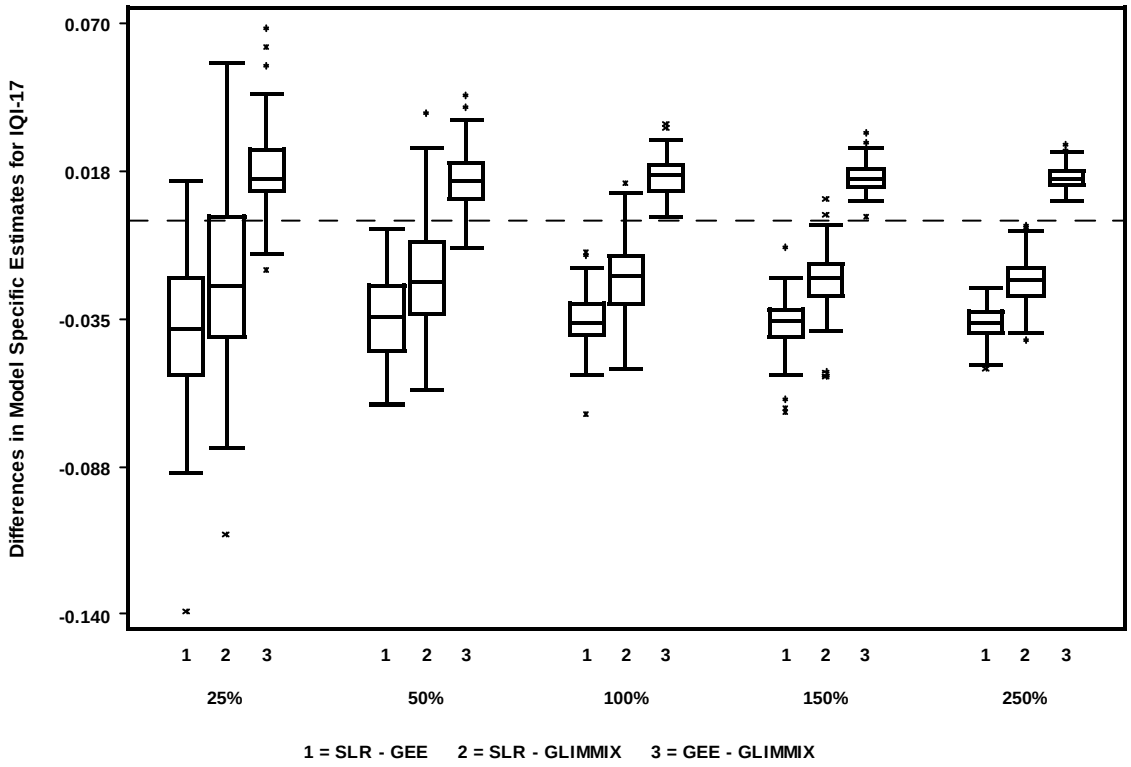


Figure 3.4.2d: Box Plots of the Differences in Model Specific Estimates of the Intercepts from a Specific Pair of Time-Trend Models Fitted to 100 Boot Strap Sampling Runs with Target Sampling at 25%, 50%, 100%, 150%, and 250% of the Observed NIS Sample Data for IQI-17.

Table 3.4.2c. Wald Test Statistics and (P-Value) Comparing Models fit to IQI-17 (IQI 17 – Acute Stroke Mortality Rate)

Test	Percent of Observed Sample	P-values					
		Cross-Sectional Models			Time-Trend Models		
		SLR vs. GEE	SLR vs. GLIM	GEE vs. GLIM	SLR vs. GEE	SLR vs. GLIM	GEE vs. GLIM
Global Wald Test	25%	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
	50%	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
	100%	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
	150%	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
	250%	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
T-Test for Intercept Difference	25%	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
	50%	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
	100%	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
	150%	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
	250%	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001

3.5 IQI 19 – Hip Fracture Mortality Rate

3.5.1 Model Parameter Estimates

Below we provide the model parameter estimates from fitting the simple logistic regression, generalized estimating equations, and generalized linear mixed model to the 2001-2003 Nationwide Inpatient Sample for IQI 19 (Hip Fracture Mortality Rate). Table 3.5.1a provides the parameter estimates associated with the cross-sectional model, and Table 3.5.1b provides the parameter estimates associate with the model that adjusts for a simple linear trend over time. Across these two tables, we see that the parameter estimates and associated standard errors are quite comparable among the three modeling approaches. In fact, the estimated correlation coefficient from the GEE modeling approach is nearly zero (ρ=0.0017) in both the cross-sectional and time-trend adjusted models. The estimated variance components associated with provider-specific random effects from the GLIMMIX model were subtle, yet statistically significant in both models ($\sigma^2_{\text{Btw Hosp}}$ = 0.126 for the cross sectional model, and 0.109 for the time-trend model). The variance component that captures provider-specific variation in the time-trend slope was statistically significant (σ^2_{Year} = 0.043).

Table 3.5.1c provides the Wald Statistics to determine whether there are statistically significant differences between the vector of parameter estimates generated by each modeling approach. The Wald Statistics consider pair-wise comparisons, and suggest that there were no significant differences between the different modeling approaches for IQI 19.

Table 3.5.1a Parameter Estimates from Cross Sectional Models fit to IQI-19 (Hip Fracture Mortality Rate)

Parameter	Simple Logistic Regression Model			Generalized Estimating Equations Model			Generalized Linear Mixed Model		
	Estimate	Std Err	p value	Estimate	Std Err	p value	Estimate	Std Err	p value
Intercept	-1.569	0.139	0.000	-1.565	0.139	0.000	-1.560	0.141	0.000
SEX	-0.310	0.043	0.000	-0.310	0.043	0.000	-0.317	0.044	0.000
AGE3	-2.610	1.007	0.010	-2.569	0.979	0.009	-2.609	1.008	0.010
AGE4	-1.244	0.459	0.007	-1.224	0.452	0.007	-1.233	0.461	0.007
AGE5	-2.137	0.587	0.000	-2.131	0.580	0.000	-2.167	0.588	0.000
AGE6	-0.905	0.271	0.001	-0.912	0.269	0.001	-0.955	0.273	0.000
AGE7	-0.729	0.224	0.001	-0.728	0.222	0.001	-0.732	0.225	0.001
AGE8	-0.860	0.225	0.000	-0.853	0.223	0.000	-0.866	0.226	0.000
AGE9	-0.525	0.170	0.002	-0.524	0.169	0.002	-0.546	0.172	0.001
AGE10	-0.453	0.122	0.000	-0.454	0.122	0.000	-0.462	0.123	0.000
AGE11	-0.713	0.096	0.000	-0.711	0.096	0.000	-0.721	0.097	0.000
AGE12	-0.449	0.071	0.000	-0.448	0.071	0.000	-0.453	0.072	0.000
AGE13	-0.296	0.061	0.000	-0.293	0.061	0.000	-0.293	0.062	0.000
AGE17	1.740	1.245	0.162	1.703	1.219	0.162	1.744	1.248	0.162
AGE18	-0.155	1.109	0.889	-0.167	1.094	0.879	-0.173	1.110	0.876
AGE19	0.950	0.835	0.256	0.951	0.826	0.249	0.965	0.837	0.249
AGE20	0.687	0.420	0.102	0.690	0.418	0.099	0.746	0.422	0.077
AGE21	-0.103	0.376	0.785	-0.098	0.373	0.793	-0.099	0.378	0.794
AGE22	0.437	0.310	0.159	0.430	0.307	0.161	0.446	0.311	0.152
AGE23	0.345	0.232	0.137	0.348	0.230	0.131	0.376	0.234	0.108
AGE24	-0.053	0.171	0.755	-0.048	0.170	0.776	-0.046	0.172	0.788
AGE25	0.141	0.124	0.256	0.140	0.123	0.255	0.135	0.125	0.279
AGE26	-0.078	0.093	0.401	-0.074	0.092	0.420	-0.078	0.093	0.405
AGE27	-0.074	0.077	0.339	-0.073	0.077	0.340	-0.078	0.078	0.313
C1	-3.211	0.166	0.000	-3.191	0.166	0.000	-3.223	0.167	0.000
C2	-2.221	0.145	0.000	-2.204	0.145	0.000	-2.224	0.146	0.000
C3	-0.330	0.142	0.020	-0.325	0.142	0.022	-0.325	0.143	0.023
C4	1.293	0.145	0.000	1.298	0.145	0.000	1.328	0.146	0.000
C5	-3.925	0.169	0.000	-3.886	0.169	0.000	-3.930	0.170	0.000
C6	-2.462	0.143	0.000	-2.440	0.143	0.000	-2.461	0.144	0.000
C7	-0.377	0.140	0.007	-0.369	0.140	0.008	-0.365	0.141	0.010
C8	1.427	0.143	0.000	1.431	0.143	0.000	1.458	0.145	0.000
C9	-1.960	0.177	0.000	-1.928	0.175	0.000	-1.920	0.178	0.000
C10	-0.639	0.146	0.000	-0.623	0.145	0.000	-0.597	0.147	0.000
C11	1.058	0.145	0.000	1.063	0.145	0.000	1.097	0.147	0.000
C12	2.582	0.161	0.000	2.584	0.161	0.000	2.620	0.163	0.000
C13	-0.502	0.152	0.001	-0.492	0.152	0.001	-0.483	0.153	0.002
ρ				0.0017	.	.			
σ^2_{Hosp}							0.126	0.017	.

The effect of the *YEAR* parameter (which captures the trend over time) was highly significant for all three modeling approaches, as seen in Table 3.5.1b below.

Table 3.5.1b Parameter Estimates from Time Trend Models fit to IQI-19 (Hip Fracture Mortality Rate)

Parameter	Simple Logistic Regression Model			Generalized Estimating Equations Model			Generalized Linear Mixed Model		
	Estimate	Std Err	p value	Estimate	Std Err	p value	Estimate	Std Err	p value
Intercept	-1.570	0.139	0.000	-1.567	0.139	0.000	-1.566	0.141	0.000
SEX	-0.311	0.043	0.000	-0.311	0.043	0.000	-0.318	0.044	0.000
AGE3	-2.609	1.007	0.010	-2.570	0.980	0.009	-2.618	1.008	0.009
AGE4	-1.242	0.460	0.007	-1.222	0.452	0.007	-1.230	0.461	0.008
AGE5	-2.141	0.587	0.000	-2.135	0.581	0.000	-2.177	0.588	0.000
AGE6	-0.897	0.270	0.001	-0.903	0.269	0.001	-0.949	0.273	0.001
AGE7	-0.727	0.224	0.001	-0.725	0.222	0.001	-0.731	0.225	0.001
AGE8	-0.858	0.225	0.000	-0.851	0.223	0.000	-0.866	0.226	0.000
AGE9	-0.521	0.170	0.002	-0.520	0.169	0.002	-0.546	0.172	0.001
AGE10	-0.448	0.122	0.000	-0.449	0.122	0.000	-0.458	0.123	0.000
AGE11	-0.714	0.096	0.000	-0.713	0.096	0.000	-0.722	0.097	0.000
AGE12	-0.447	0.071	0.000	-0.446	0.071	0.000	-0.454	0.072	0.000
AGE13	-0.296	0.061	0.000	-0.293	0.061	0.000	-0.293	0.062	0.000
AGE17	1.751	1.246	0.160	1.715	1.220	0.160	1.770	1.248	0.156
AGE18	-0.159	1.109	0.886	-0.170	1.093	0.876	-0.170	1.110	0.879
AGE19	0.952	0.836	0.254	0.954	0.826	0.248	0.975	0.837	0.244
AGE20	0.682	0.420	0.105	0.685	0.418	0.101	0.747	0.422	0.077
AGE21	-0.107	0.376	0.777	-0.102	0.373	0.785	-0.099	0.378	0.794
AGE22	0.443	0.310	0.152	0.437	0.307	0.154	0.452	0.311	0.146
AGE23	0.347	0.232	0.135	0.350	0.230	0.129	0.380	0.234	0.105
AGE24	-0.059	0.171	0.729	-0.054	0.170	0.750	-0.049	0.172	0.778
AGE25	0.143	0.124	0.249	0.142	0.123	0.249	0.137	0.125	0.272
AGE26	-0.079	0.093	0.393	-0.075	0.092	0.413	-0.077	0.093	0.408
AGE27	-0.074	0.077	0.339	-0.073	0.077	0.340	-0.077	0.078	0.320
C1	-3.213	0.166	0.000	-3.192	0.166	0.000	-3.221	0.168	0.000
C2	-2.220	0.145	0.000	-2.204	0.145	0.000	-2.219	0.146	0.000
C3	-0.326	0.142	0.022	-0.321	0.142	0.024	-0.317	0.143	0.027
C4	1.299	0.145	0.000	1.304	0.145	0.000	1.338	0.146	0.000
C5	-3.927	0.169	0.000	-3.887	0.169	0.000	-3.928	0.170	0.000
C6	-2.461	0.143	0.000	-2.439	0.143	0.000	-2.456	0.144	0.000
C7	-0.374	0.140	0.008	-0.366	0.140	0.009	-0.358	0.142	0.012
C8	1.433	0.143	0.000	1.437	0.143	0.000	1.468	0.145	0.000
C9	-1.966	0.177	0.000	-1.933	0.175	0.000	-1.919	0.178	0.000
C10	-0.641	0.146	0.000	-0.625	0.145	0.000	-0.592	0.147	0.000
C11	1.057	0.145	0.000	1.062	0.145	0.000	1.102	0.147	0.000
C12	2.589	0.161	0.000	2.592	0.161	0.000	2.634	0.163	0.000
C13	-0.501	0.152	0.001	-0.490	0.152	0.001	-0.478	0.154	0.002
YEAR	-0.063	0.018	0.001	-0.066	0.019	0.000	-0.066	0.021	0.002
ρ				0.0017	.	.			
σ^2_{Hosp}							0.109	0.018	.
σ^2_{Year}							0.043	0.019	.

Table 3.5.1c Wald Test Statistics and (P-Value) Comparing Models fit to IQI-19 (Hip Fracture Mortality Rate)

	Cross Sectional Model			Time Trend Model		
	SLR	**GEE**	**GLIMMIX**	**SLR**	**GEE**	**GLIMMIX**
SLR		1.473 (1.000)	3.396 (1.000)		1.557 (1.000)	3.440 (1.000)
GEE	1.178 (1.000)		2.391 (1.000)	1.252 (1.000)		2.509 (1.000)
GLIMMIX	2.614 (1.000)	2.332 (1.000)		2.696 (1.000)	2.470 (1.000)	

- Wald Test uses the estimated covariance matrix from the model listed in each row

Table 3.5.1d below provides the mean and standard deviation of differences between model predictions (expected rates above the diagonal, and risk-adjusted rates below the diagonal) from a random sample of 50 providers within the NIS reference population for IQI 17. The mean difference in expected rates and risk-adjusted rates was nearly zero for all modeling approach pairwise comparisons – despite the fact that the GLIMMIX modeling approaches identified statistically significant variance components associated with the random effects.

Table 3.5.1d Estimated Differences (and Standard Deviation) in Provider-Level Model Predictions of Expected and Risk Adjusted Rates for IQI-19 (Hip Fracture Mortality Rate)

	Cross Sectional Model			Time Trend Model		
	SLR	**GEE**	**GLIMMIX**	**SLR**	**GEE**	**GLIMMIX**
SLR		-0.000 (0.000)	-0.000 (0.001)		-0.000 (0.000)	-0.000 (0.001)
GEE	0.004 (0.006)		0.000 (0.000)	0.004 (0.006)		0.000 (0.000)
GLIMMIX	0.001 (0.005)	-0.003 (0.005)		0.001 (0.005)	-0.003 (0.006)	

***** In each 3x3 table above, Expected Rate Differences (and Standard Deviations) are above the diagonal, and Adjusted Rate Differences (and Standard Deviations) are below the diagonal.

3.5.2 Model Parameter Estimates from Fitting Models to Boot Strap Samples of NIS Data

Both cross sectional and time-trend adjusted models (especially GLIMMIX) fitted for quality indicators IQI-19 (Abdominal Aortic Aneurysm Repair Mortality Rate) for bootstrap sampling runs with target sample size of 25% and 50% of the observed NIS sample had convergence problems due to small sample size or rare occurrence of the event.

Tables 3.5.2a and 3.5.2b provide an overview of the results for the cross sectional and time-trend adjusted models fit to 100 bootstrap samples with population size approximately 100% of the size of the NIS sample. Each table provides the mean across

the 100 bootstrap samples of the model parameter estimates and standard errors from the Simple Logistic Regression Model. The table also provides average absolute and relative differences between each pair of modeling approaches. Box plots of model specific intercept estimates and differences in intercept estimates between each pair of modeling approaches when target sample sizes ranging from 100% to 250% of the observed NIS samples for cross-sectional models are displayed in Figures 3.5.2a and 3.5.2c. Figures 3.5.2b and 3.5.2d presents the similar boxplots for the time-trend models. The results (p-values) from global Wald tests performed to assess whether there are statistically significant difference between a pair of model parameter estimates are provided in Table 3.5.2c. Results (p-values) from pair-wise t-test to test for intercept differences are also included in Table 3.5.2c. Both Wald and t-tests are performed at 5% level of significance.

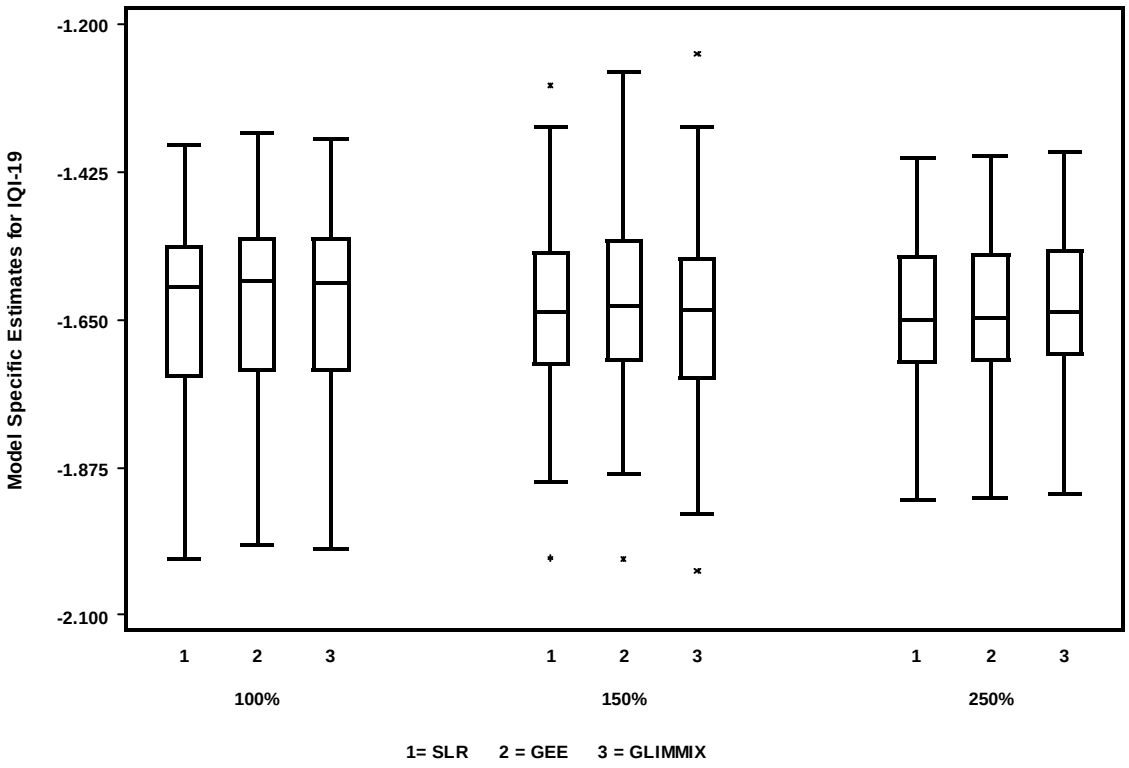
Both cross-sectional and time-trend model parameter estimates from three modeling approaches match very closely. Larger differences and relative differences among any pair of models are found for parameters Age25 and Age 26 compared to other parameters. Box plots of intercept estimates for all three modeling approaches show similar spread for both cross-sectional and time-trend models. Box plots of intercept differences show that the results of SLR and GEE modeling approaches are more similar to each other compared to GLIMMIX. Variability in the intercept differences from valid bootstrap runs is smaller for target sample size at 50% as expected. All overall Wald test results show significant differences in the three modeling approaches for both cross-sectional and time-trend models at all target sample sizes explored for IQI-19. Intercept differences are statistically significant for SLR and GEE for both cross-sectional and time-trend models based on the paired t-test results. Table 3.5.2c shows that there are no statistically significant differences in intercept differences found between SLR and GLIMMIX models for the cross-sectional model, and between the GEE and GLIMMIX models for time-trend model at 100% target sample size.

Table 3.5.2a. Average Models Estimates and Relative Standard Errors from Simple Logistic Model Fit, Differences and Relative Differences in Model Estimates from a Specific Pair of Modeling Approaches for the Inpatient Quality Indicator IQI-19 (IQI-19 – Hip Fracture Mortality Rate) For Each Parameter in the Model (Cross-Sectional)

Cross-Sectional								
Parameter	Simple Logistic Regression		Difference between model estimates from a pair of modeling approaches			Relative Difference between model estimates from a pair of modeling approaches		
	Estimate	Std. Error	SLR vs. GEE	SLR vs. GLIM	GEE vs. GLIM	SLR vs. GEE	SLR vs. GLIM	GEE vs. GLIM
Intercept	-1.634	0.1420	-0.011	-0.006	0.005	0.007	0.004	0.003
SEX	-0.310	0.0434	-0.000	0.018	0.018	0.001	0.056	0.057
AGE3	-2.386	0.8902	-0.092	-0.005	0.087	0.039	0.002	0.037
AGE4	-1.431	0.5252	-0.049	0.011	0.060	0.035	0.007	0.042
AGE5	-2.194	0.6436	-0.024	0.064	0.088	0.011	0.029	0.040
AGE6	-1.080	0.2989	0.011	0.091	0.080	0.010	0.081	0.071
AGE7	-0.765	0.2298	-0.014	-0.001	0.013	0.019	0.001	0.018
AGE8	-0.836	0.2263	-0.017	0.015	0.032	0.021	0.018	0.039
AGE9	-0.583	0.1737	-0.009	0.036	0.045	0.015	0.061	0.076
AGE10	-0.481	0.1232	-0.002	0.019	0.021	0.005	0.039	0.044
AGE11	-0.660	0.0952	-0.005	0.018	0.023	0.008	0.027	0.035
AGE12	-0.464	0.0718	-0.004	0.012	0.016	0.009	0.025	0.034
AGE13	-0.269	0.0608	-0.008	-0.005	0.003	0.030	0.017	0.012
AGE17	1.535	1.2103	0.081	-0.021	-0.102	0.054	0.014	0.068
AGE18	0.174	1.0819	0.004	-0.031	-0.035	0.023	0.161	0.184
AGE19	0.872	0.9449	-0.019	-0.052	-0.033	0.021	0.058	0.036
AGE20	0.895	0.4421	-0.010	-0.117	-0.107	0.011	0.123	0.111
AGE21	-0.422	0.4326	-0.029	-0.020	0.008	0.070	0.049	0.021
AGE22	0.403	0.3126	0.012	-0.019	-0.031	0.030	0.047	0.076
AGE23	0.378	0.2366	0.002	-0.058	-0.060	0.004	0.143	0.148
AGE25	-0.013	0.1725	-0.008	-0.010	-0.002	0.954	1.373	0.623
AGE24	0.101	0.1233	0.002	0.012	0.009	0.024	0.125	0.100
AGE26	-0.019	0.0926	-0.007	-0.002	0.005	0.449	0.097	0.356
AGE27	-0.100	0.0771	-0.002	0.006	0.008	0.016	0.060	0.076
C1	-3.130	0.1681	-0.051	0.040	0.091	0.016	0.013	0.029
C2	-2.159	0.1474	-0.038	0.022	0.060	0.018	0.010	0.028
C3	-0.285	0.1446	-0.011	0.001	0.012	0.040	0.003	0.043
C4	1.359	0.1473	-0.010	-0.066	-0.056	0.007	0.047	0.040
C5	-3.875	0.1716	-0.091	0.033	0.124	0.024	0.008	0.032
C6	-2.418	0.1452	-0.051	0.019	0.070	0.021	0.008	0.029
C7	-0.316	0.1429	-0.015	-0.007	0.008	0.049	0.023	0.026
C8	1.477	0.1457	-0.006	-0.061	-0.055	0.004	0.040	0.036
C9	-1.925	0.1820	-0.084	-0.054	0.030	0.045	0.029	0.016
C10	-0.551	0.1484	-0.034	-0.054	-0.020	0.065	0.104	0.039
C11	1.122	0.1480	-0.008	-0.066	-0.058	0.007	0.057	0.050
C12	2.685	0.1637	-0.001	-0.086	-0.085	0.000	0.031	0.031
C13	-0.440	0.1545	-0.022	-0.019	0.003	0.050	0.044	0.007

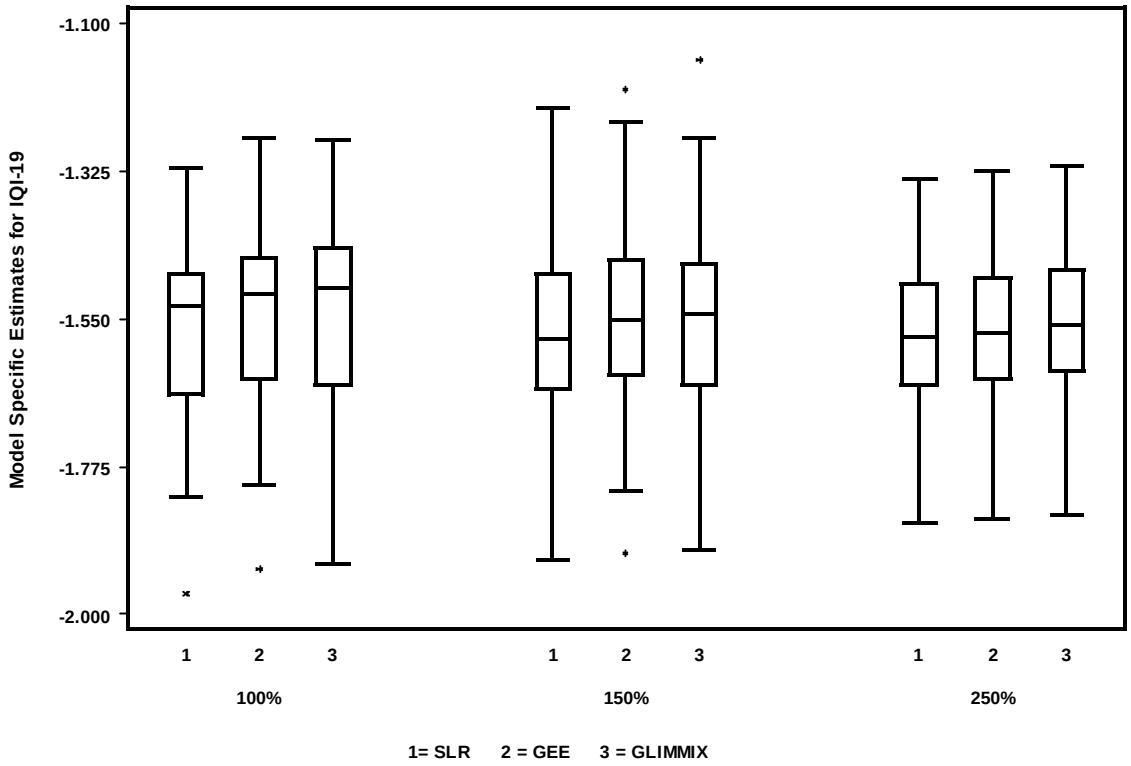
Table 3.5.2b. Average Models Estimates and Relative Standard Errors from Simple Logistic Model Fit, Differences and Relative Differences in Model Estimates from a Specific Pair of Modeling Approaches for the Inpatient Quality Indicator IQI-19 (IQI-19 – Hip Fracture Mortality Rate) For Each Parameter in the Model (Time-Trend)

Time-Trend								
Parameter	Simple Logistic Regression		Difference between model estimates from a pair of modeling approaches			Relative Difference between model estimates from a pair of modeling approaches		
	Estimate	Std. Error	SLR vs. GEE	SLR vs. GLIM	GEE vs. GLIM	SLR vs. GEE	SLR vs. GLIM	GEE vs. GLIM
Intercept	-1.571	0.1431	-0.021	-0.022	-0.001	0.013	0.014	0.001
SEX	-0.311	0.0434	0.000	0.018	0.018	0.000	0.057	0.057
AGE3	-2.389	0.8905	-0.089	-0.003	0.086	0.038	0.001	0.037
AGE4	-1.428	0.5253	-0.050	0.011	0.061	0.035	0.008	0.043
AGE5	-2.198	0.6439	-0.023	0.065	0.088	0.010	0.029	0.040
AGE6	-1.073	0.2989	0.010	0.090	0.080	0.009	0.081	0.072
AGE7	-0.764	0.2298	-0.015	-0.001	0.014	0.020	0.002	0.018
AGE8	-0.834	0.2263	-0.016	0.017	0.033	0.020	0.020	0.040
AGE9	-0.579	0.1737	-0.008	0.037	0.046	0.014	0.063	0.077
AGE10	-0.476	0.1232	-0.002	0.022	0.024	0.004	0.044	0.048
AGE11	-0.662	0.0952	-0.005	0.018	0.023	0.007	0.027	0.035
AGE12	-0.462	0.0718	-0.003	0.014	0.017	0.007	0.030	0.037
AGE13	-0.269	0.0608	-0.008	-0.004	0.004	0.030	0.016	0.014
AGE17	1.553	1.2106	0.076	-0.023	-0.099	0.050	0.015	0.065
AGE18	0.167	1.0820	0.000	-0.032	-0.032	0.003	0.175	0.178
AGE19	0.874	0.9453	-0.021	-0.055	-0.034	0.024	0.061	0.037
AGE20	0.894	0.4420	-0.009	-0.117	-0.108	0.010	0.123	0.113
AGE21	-0.424	0.4327	-0.029	-0.021	0.007	0.070	0.051	0.019
AGE22	0.410	0.3124	0.011	-0.020	-0.031	0.027	0.047	0.075
AGE23	0.382	0.2366	0.001	-0.058	-0.059	0.004	0.141	0.144
AGE24	-0.018	0.1725	-0.008	-0.014	-0.006	0.586	1.248	0.810
AGE25	0.103	0.1233	0.003	0.012	0.010	0.026	0.127	0.101
AGE26	-0.020	0.0926	-0.008	-0.004	0.003	0.495	0.251	0.252
AGE27	-0.100	0.0771	-0.002	0.006	0.007	0.016	0.056	0.071
C1	-3.132	0.1681	-0.053	0.038	0.091	0.017	0.012	0.029
C2	-2.159	0.1474	-0.038	0.021	0.059	0.018	0.010	0.027
C3	-0.281	0.1446	-0.011	0.001	0.012	0.041	0.003	0.044
C4	1.365	0.1473	-0.010	-0.066	-0.056	0.008	0.047	0.040
C5	-3.878	0.1716	-0.093	0.032	0.125	0.024	0.008	0.033
C6	-2.417	0.1452	-0.051	0.018	0.069	0.021	0.007	0.029
C7	-0.313	0.1429	-0.015	-0.008	0.008	0.050	0.025	0.025
C8	1.483	0.1458	-0.007	-0.061	-0.054	0.005	0.040	0.036
C9	-1.931	0.1820	-0.086	-0.058	0.028	0.046	0.031	0.015
C10	-0.551	0.1484	-0.035	-0.057	-0.022	0.065	0.108	0.043
C11	1.122	0.1480	-0.008	-0.068	-0.059	0.008	0.059	0.051
C12	2.693	0.1638	-0.002	-0.086	-0.084	0.001	0.031	0.031
C13	-0.439	0.1545	-0.022	-0.020	0.002	0.051	0.047	0.004
Year	-0.063	0.0180	0.010	0.017	0.007	0.148	0.237	0.090



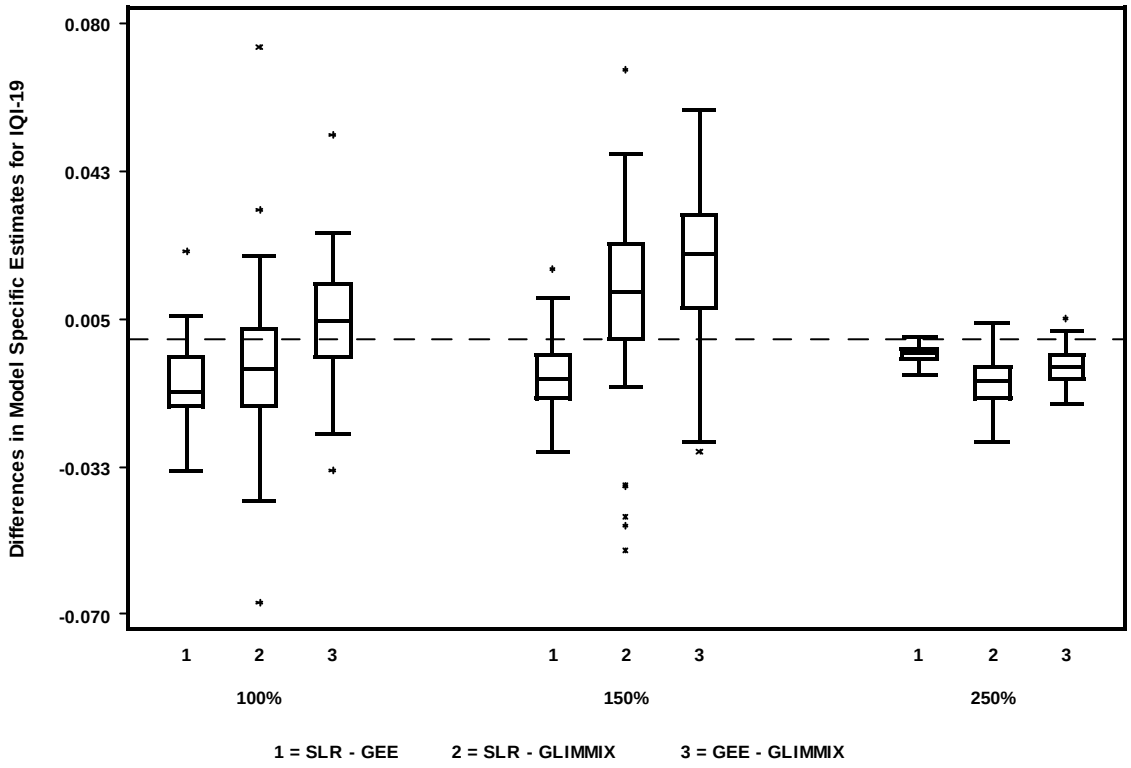
Note: Box plots for target sample at 25% and 50% are not included since the models did not converge**.**

Figure 3.5.2a: Box Plots of the Model Specific Estimates of the Intercepts Obtained from fitting Cross-Sectional Models to 100 Boot Strap Sampling with Target Sampling at 25%, 50%, 100%, 150%, and 250% of the Observed NIS Sample Data for IQI-19.



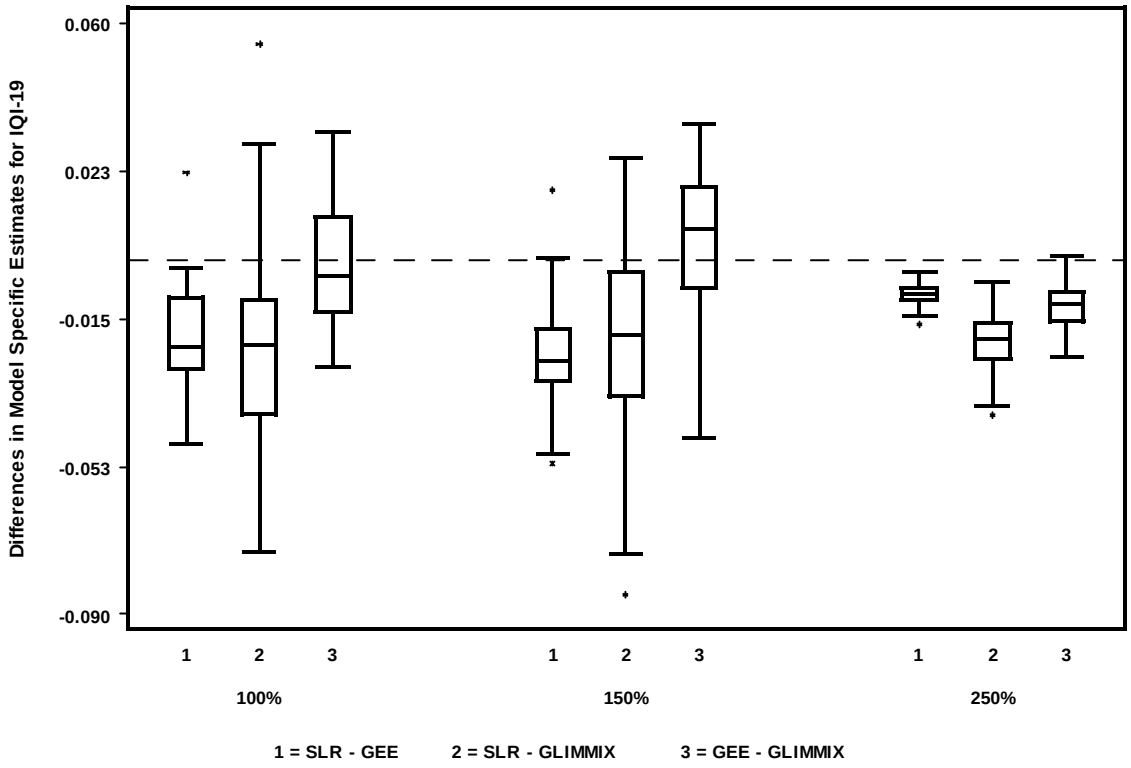
Note: Box plots for target sample at 25% and 50% are not included since the models did not converge**.**

Figure 3.5.2b: Box Plots of the Model Specific Estimates of the Intercepts Obtained from fitting Time-Trend Models to 100 Boot Strap Sampling with Target Sampling at 25%, 50%, 100%, 150%, and 250% of the Observed NIS Sample Data for IQI-19.



Note: Box plots for target sample at 25% and 50% are not included since the models did not converge**.**

Figure 3.5.2c: Box Plots of the Differences in Model Specific Estimates of the Intercepts from a Specific Pair of Cross-Sectional Models Fitted to 100 Boot Strap Sampling Runs with Target Sampling at 25%, 50%, 100%, 150%, and 250% of the Observed NIS Sample Data for IQI-19.



Note: Box plots for target sample at 25% and 50% are not included since the models did not converge**.**

Figure 3.5.2d: Box Plots of the Differences in Model Specific Estimates of the Intercepts from a Specific Pair of Time-Trend Models Fitted to 100 Boot Strap Sampling Runs with Target Sampling at 25%, 50%, 100%, 150%, and 250% of the Observed NIS Sample Data for IQI-19.

Table 3.5.2c Wald Test Statistics and (P-Value) Comparing Models fit to IQI-19 (IQI-19 – Hip Fracture Mortality Rate)

Test	Percent of Observed Sample	P-values					
		Cross-Sectional Models			Time-Trend Models		
		SLR vs. GEE	SLR vs. GLIM	GEE vs. GLIM	SLR vs. GEE	SLR vs. GLIM	GEE vs. GLIM
Global Wald Test	25%	.	.	.	.	.	.
	50%	.	.	.	.	.	.
	100%	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
	150%	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
	250%	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
T-Test for Intercept Difference	25%						
	50%						
	100%	< 0.0001	0.1598	0.0895	< 0.0001	< 0.0001	0.6859
	150%	< 0.0001	0.0026	< 0.0001	< 0.0001	< 0.0001	0.0842
	250%	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001

3.6 IQI 25 – Bilateral Cardiac Catheterization Rate

3.6.1 Model Parameter Estimates

Below we provide the model parameter estimates from fitting the simple logistic regression, generalized estimating equations, and generalized linear mixed model to the 2001-2003 Nationwide Inpatient Sample for IQI 25 (Bilateral Cardiac Catheterization Rate). Table 3.6.1a provides the parameter estimates associated with the cross-sectional model, and Table 3.6.1b provides the parameter estimates associate with the model that adjusts for a simple linear trend over time. Across these two tables, we see that the parameter estimates and associated standard errors are comparable, but with noticeable differences among the three modeling approaches. The estimated correlation coefficient from the GEE modeling approach was largest for this Quality Indicator (compared to the other four QIs investigated), with ρ=0.0412 in the cross-sectional and ρ=0.0405 in the time-trend adjusted model. The estimated variance components associated with provider-specific random effects from the GLIMMIX model were also comparatively larger, and statistically significant in both models ($\sigma^2_{\text{Btw Hosp}} = 0.779$ for the cross sectional model, and 0.766 for the time-trend model). The variance component that captures provider-specific variation in the time-trend slope was statistically significant ($\sigma^2_{\text{Year}} = 0.031$).

The effect of the *YEAR* parameter (which captures the trend over time) was highly significant for all three modeling approaches, as seen in Table 3.6.1b below.

Table 3.6.1c provides the Wald Statistics to determine whether there are statistically significant differences between the vector of parameter estimates generated by each modeling approach. The Wald Statistics consider pair-wise comparisons, and suggest that there were highly significant differences between the different modeling approaches for IQI 25.

Table 3.6.1a Parameter Estimates from Cross Sectional Models fit to IQI-25 (Bilateral Cardiac Catheterization Rate)

Parameter	Simple Logistic Regression Model			Generalized Estimating Equations Model			Generalized Linear Mixed Model		
	Estimate	Std Err	p value	Estimate	Std Err	p value	Estimate	Std Err	p value
Intercept	-1.820	0.038	0.000	-1.871	0.042	0.000	-2.169	0.050	0.000
SEX	-0.115	0.036	0.001	-0.121	0.036	0.001	-0.127	0.038	0.001
AGE1	-0.307	0.214	0.152	-0.283	0.209	0.174	-0.273	0.222	0.219
AGE2	-0.556	0.158	0.000	-0.465	0.150	0.002	-0.499	0.163	0.002
AGE3	-0.829	0.089	0.000	-0.687	0.083	0.000	-0.751	0.092	0.000
AGE4	-0.964	0.061	0.000	-0.830	0.058	0.000	-0.907	0.063	0.000
AGE5	-0.903	0.043	0.000	-0.771	0.041	0.000	-0.840	0.044	0.000
AGE6	-0.850	0.036	0.000	-0.722	0.035	0.000	-0.787	0.037	0.000
AGE7	-0.765	0.033	0.000	-0.640	0.032	0.000	-0.694	0.034	0.000
AGE8	-0.591	0.031	0.000	-0.500	0.031	0.000	-0.537	0.033	0.000
AGE9	-0.472	0.031	0.000	-0.385	0.030	0.000	-0.410	0.032	0.000
AGE10	-0.308	0.030	0.000	-0.238	0.029	0.000	-0.248	0.031	0.000
AGE11	-0.212	0.030	0.000	-0.153	0.029	0.000	-0.155	0.031	0.000
AGE12	-0.072	0.030	0.015	-0.037	0.029	0.202	-0.031	0.031	0.315
AGE13	0.002	0.031	0.941	0.021	0.031	0.492	0.029	0.033	0.367
AGE15	0.555	0.366	0.129	0.469	0.364	0.197	0.462	0.385	0.230
AGE16	0.473	0.259	0.067	0.456	0.247	0.065	0.492	0.267	0.066
AGE17	0.251	0.165	0.128	0.239	0.154	0.120	0.268	0.170	0.116
AGE18	0.277	0.104	0.008	0.283	0.097	0.004	0.302	0.107	0.005
AGE19	0.278	0.070	0.000	0.297	0.065	0.000	0.323	0.072	0.000
AGE20	0.299	0.056	0.000	0.295	0.053	0.000	0.321	0.058	0.000
AGE21	0.221	0.049	0.000	0.210	0.048	0.000	0.229	0.051	0.000
AGE22	0.126	0.046	0.006	0.148	0.045	0.001	0.160	0.048	0.001
AGE23	0.152	0.044	0.001	0.162	0.043	0.000	0.177	0.046	0.000
AGE24	0.057	0.043	0.180	0.079	0.042	0.059	0.084	0.045	0.059
AGE25	0.070	0.042	0.094	0.084	0.041	0.039	0.089	0.043	0.041
AGE26	0.034	0.041	0.408	0.055	0.040	0.173	0.058	0.043	0.175
AGE27	-0.045	0.043	0.300	-0.022	0.042	0.601	-0.028	0.045	0.543
C1	-1.333	0.063	0.000	-1.119	0.058	0.000	-1.263	0.064	0.000
C2	-0.980	0.034	0.000	-0.822	0.034	0.000	-0.916	0.036	0.000
C3	-0.123	0.033	0.000	-0.079	0.032	0.014	-0.088	0.034	0.011
C4	0.218	0.041	0.000	0.275	0.039	0.000	0.298	0.043	0.000
C5	-1.340	0.032	0.000	-1.191	0.035	0.000	-1.319	0.034	0.000
C6	-0.658	0.031	0.000	-0.576	0.030	0.000	-0.634	0.032	0.000
C7	0.159	0.035	0.000	0.169	0.034	0.000	0.177	0.036	0.000
C8	0.548	0.054	0.000	0.544	0.052	0.000	0.570	0.056	0.000
C9	-1.019	0.049	0.000	-0.939	0.048	0.000	-1.033	0.050	0.000
C10	-0.451	0.035	0.000	-0.379	0.034	0.000	-0.415	0.036	0.000
C11	0.268	0.037	0.000	0.307	0.036	0.000	0.343	0.038	0.000
C12	0.836	0.041	0.000	0.860	0.040	0.000	0.947	0.043	0.000
C13	0.638	0.039	0.000	0.625	0.038	0.000	0.681	0.041	0.000
C14	0.491	0.037	0.000	0.502	0.036	0.000	0.555	0.038	0.000
C15	0.921	0.031	0.000	0.907	0.031	0.000	0.995	0.032	0.000
C16	1.032	0.053	0.000	1.028	0.052	0.000	1.128	0.055	0.000
C17	-0.954	0.032	0.000	-0.933	0.033	0.000	-1.014	0.033	0.000
C18	-0.308	0.031	0.000	-0.282	0.030	0.000	-0.309	0.032	0.000
C19	0.214	0.036	0.000	0.202	0.036	0.000	0.219	0.038	0.000
C20	0.628	0.080	0.000	0.612	0.079	0.000	0.652	0.084	0.000
C21	-0.068	0.028	0.017	-0.008	0.028	0.761	-0.010	0.030	0.738
ρ				0.0412	.	.			
σ^2_{Hosp}							0.779	0.041	.

Table 3.6.1b Parameter Estimates from Time Trend Models fit to IQI-25 (Bilateral Cardiac Catheterization Rate)

Parameter	Simple Logistic Regression Model			Generalized Estimating Equations Model			Generalized Linear Mixed Model		
	Estimate	Std Err	p value	Estimate	Std Err	p value	Estimate	Std Err	p value
Intercept	-1.817	0.038	0.000	-1.870	0.042	0.000	-2.167	0.050	0.000
SEX	-0.116	0.036	0.001	-0.122	0.036	0.001	-0.130	0.038	0.001
AGE1	-0.298	0.215	0.165	-0.277	0.209	0.184	-0.263	0.222	0.236
AGE2	-0.549	0.158	0.001	-0.464	0.150	0.002	-0.499	0.163	0.002
AGE3	-0.824	0.089	0.000	-0.688	0.083	0.000	-0.755	0.092	0.000
AGE4	-0.963	0.061	0.000	-0.832	0.058	0.000	-0.908	0.063	0.000
AGE5	-0.901	0.043	0.000	-0.773	0.041	0.000	-0.844	0.044	0.000
AGE6	-0.847	0.036	0.000	-0.723	0.035	0.000	-0.788	0.037	0.000
AGE7	-0.764	0.033	0.000	-0.642	0.032	0.000	-0.696	0.034	0.000
AGE8	-0.589	0.031	0.000	-0.500	0.031	0.000	-0.539	0.033	0.000
AGE9	-0.470	0.031	0.000	-0.386	0.030	0.000	-0.411	0.032	0.000
AGE10	-0.308	0.030	0.000	-0.240	0.029	0.000	-0.250	0.031	0.000
AGE11	-0.213	0.030	0.000	-0.154	0.029	0.000	-0.157	0.031	0.000
AGE12	-0.072	0.030	0.015	-0.038	0.029	0.190	-0.032	0.031	0.298
AGE13	0.003	0.031	0.935	0.020	0.031	0.503	0.028	0.033	0.384
AGE15	0.568	0.366	0.121	0.475	0.365	0.192	0.462	0.385	0.230
AGE16	0.475	0.259	0.066	0.458	0.248	0.064	0.504	0.267	0.060
AGE17	0.250	0.165	0.130	0.239	0.154	0.121	0.275	0.170	0.106
AGE18	0.279	0.104	0.008	0.286	0.097	0.003	0.301	0.108	0.005
AGE19	0.281	0.070	0.000	0.298	0.065	0.000	0.328	0.072	0.000
AGE20	0.301	0.056	0.000	0.297	0.054	0.000	0.323	0.058	0.000
AGE21	0.222	0.049	0.000	0.211	0.048	0.000	0.232	0.051	0.000
AGE22	0.127	0.046	0.006	0.149	0.045	0.001	0.163	0.048	0.001
AGE23	0.154	0.044	0.001	0.163	0.043	0.000	0.179	0.046	0.000
AGE24	0.058	0.043	0.175	0.080	0.042	0.056	0.086	0.045	0.053
AGE25	0.070	0.042	0.090	0.084	0.041	0.038	0.091	0.043	0.037
AGE26	0.034	0.041	0.408	0.055	0.040	0.172	0.060	0.043	0.165
AGE27	-0.045	0.043	0.303	-0.022	0.043	0.610	-0.026	0.045	0.573
C1	-1.343	0.063	0.000	-1.122	0.058	0.000	-1.266	0.064	0.000
C2	-0.987	0.034	0.000	-0.826	0.034	0.000	-0.919	0.036	0.000
C3	-0.128	0.033	0.000	-0.081	0.032	0.011	-0.090	0.034	0.009
C4	0.215	0.041	0.000	0.273	0.040	0.000	0.298	0.043	0.000
C5	-1.345	0.032	0.000	-1.192	0.035	0.000	-1.320	0.034	0.000
C6	-0.660	0.031	0.000	-0.576	0.030	0.000	-0.632	0.032	0.000
C7	0.158	0.035	0.000	0.169	0.034	0.000	0.177	0.036	0.000
C8	0.544	0.054	0.000	0.542	0.053	0.000	0.570	0.056	0.000
C9	-1.026	0.049	0.000	-0.942	0.048	0.000	-1.035	0.050	0.000
C10	-0.459	0.035	0.000	-0.382	0.034	0.000	-0.417	0.036	0.000
C11	0.267	0.037	0.000	0.307	0.036	0.000	0.344	0.038	0.000
C12	0.836	0.041	0.000	0.861	0.040	0.000	0.949	0.043	0.000
C13	0.633	0.039	0.000	0.623	0.039	0.000	0.682	0.041	0.000
C14	0.489	0.037	0.000	0.501	0.036	0.000	0.555	0.038	0.000
C15	0.920	0.031	0.000	0.907	0.031	0.000	0.995	0.032	0.000
C16	1.034	0.053	0.000	1.030	0.052	0.000	1.129	0.055	0.000
C17	-0.962	0.032	0.000	-0.936	0.033	0.000	-1.017	0.033	0.000
C18	-0.313	0.031	0.000	-0.285	0.030	0.000	-0.311	0.032	0.000
C19	0.212	0.036	0.000	0.201	0.036	0.000	0.219	0.038	0.000
C20	0.626	0.080	0.000	0.612	0.079	0.000	0.648	0.084	0.000
C21	-0.069	0.028	0.014	-0.009	0.028	0.755	-0.009	0.030	0.753
YEAR	-0.073	0.005	0.000	-0.072	0.007	0.000	-0.077	0.016	0.000
ρ				0.0405	.	.			
σ^2_{Hosp}							0.766	0.042	.
σ^2_{Year}							0.031	0.007	

Table 3.6.1c Wald Test Statistics and (P-Value) Comparing Models fit to IQI-25 (Bilateral Cardiac Catheterization Rate)

	Cross Sectional Model			Time Trend Model		
	SLR	**GEE**	**GLIMMIX**	**SLR**	**GEE**	**GLIMMIX**
SLR		466.78 (0.000)	4037.6 (0.000)		457.86 (0.000)	4035.7 (0.000)
GEE	146.63 (0.000)		213.69 (0.000)	145.58 (0.000)		214.20 (0.000)
GLIMMIX	158.45 (0.000)	309.24 (0.000)		155.51 (0.000)	308.95 (0.000)	

- Wald Test uses the estimated covariance matrix from the model listed in each row

Table 3.6.1d below provides the mean and standard deviation of differences between model predictions (expected rates above the diagonal, and risk-adjusted rates below the diagonal) from a random sample of 50 providers within the NIS reference population for IQI 17. For example, the mean difference in expected rates between the GEE and GLIMMIX approaches for the cross sectional model was -0.016 (relative to a national mean response rate of 0.076 from Table 3.1). This estimated difference is quite high relative to the national mean response rate, demonstrating that model choice could have a significant effect on provider-level estimates.

Mean differences (and standard deviations) attributable to model specification (simple logistic vs GEE vs GLIMMIX) for the risk-adjusted rates appear to be higher than the expected rates.

Table 3.6.1d Estimated Differences (and Standard Deviation) in Provider-Level Model Predictions of Expected and Risk Adjusted Rates for IQI-25 (Bilateral Cardiac Catheterization Rate)

	Cross Sectional Model			Time Trend Model		
	SLR	**GEE**	**GLIMMIX**	**SLR**	**GEE**	**GLIMMIX**
SLR		-0.005 (0.003)	0.016 (0.006)		-0.005 (0.003)	0.016 (0.006)
GEE	0.008 (0.008)		0.021 (0.003)	0.008 (0.007)		0.021 (0.003)
GLIMMIX	-0.036 (0.044)	-0.044 (0.049)		-0.036 (0.045)	-0.044 (0.051)	

***** In each 3x3 table above, Expected Rate Differences (and Standard Deviations) are above the diagonal, and Adjusted Rate Differences (and Standard Deviations) are below the diagonal.

3.6.2 Model Parameter Estimates from Fitting Models to Boot Strap Samples of NIS Data

Both cross sectional and time-trend adjusted models fitted successfully for quality indicators IQI-25 (Abdominal Aortic Aneurysm Repair Mortality Rate) for all bootstrap sampling runs with target sample size of 25% to 250% of the observed NIS sample.

Tables 3.6.2a and 3.6.2b provide an overview of the results for the cross sectional and time-trend adjusted models fit to 100 bootstrap samples with population size approximately 100% of the size of the NIS sample. Each table provides the mean across the 100 bootstrap samples of the model parameter estimates and standard errors from the Simple Logistic Regression Model. The table also provides average absolute and relative differences between each pair of modeling approaches. Box plots of model specific intercept estimates and differences in intercept estimates between each pair of modeling approaches when target sample sizes ranging from 25% to 250% of the observed NIS samples for cross-sectional models are displayed in Figures 3.6.2a and 3.6.2c. Figures 3.6.2b and 3.6.2d presents the similar boxplots for the time-trend models. The results (p-values) from global Wald tests performed to assess whether there are statistically significant difference between a pair of model parameter estimates are provided in Table 3.6.2c. Results (p-values) from pair-wise t-test to test for intercept differences are also included in Table 3.6.2c. Both Wald and t-tests are performed at 5% level of significance.

Parameter estimates from three modeling approaches for both cross-sectional and time-trend models are comparable since the differences in model specific parameter estimates between models are so small. Relative differences among a pair of models are also small for all paramteres except for Age13 covariate as seen from Tables 3.6.2.a and 3.6.2b. Box plots of intercept estimates from three modeling approaches show that SLR and GEE estimates are closer to each other than to GLIMMIX estimates, and GLIMIX intercept estimates are consistently smaller compared to SLR and GEE for all target sample sizes considered. This is true for both cross-sectional and time-trend models. Box plots of intercept differences in SLR and GEE estimates are close to the reference line zero compared to the other two differences (SLR vs. GLIMMIX and GEE vs. GLIMMIX) which provides additional evidence that SLR and GEE intercept models are similar. All global Wald test as well as t-tests for intercept differences are found to be statistically significant for both cross-sectional and time –trend models at all target sample sizes explored for IQI-25.

Table 3.6.2a. Average Models Estimates and Relative Standard Errors from Simple Logistic Model Fit, Differences and Relative Differences in Model Estimates from a Specific Pair of Modeling Approaches for the In Patient Quality Indicator IQI-25 (IQI 25 – Bilateral Cardiac Catheterization Rate) For Each Parameter in the Model (Cross-Sectional)

Cross-Sectional								
Parameter	Simple Logistic Regression		Difference between model estimates from a pair of modeling approaches			Relative Difference between model estimates from a pair of modeling approaches		
	Estimate	Std. Error	SLR vs. GEE	SLR vs. GLIM	GEE vs. GLIM	SLR vs. GEE	SLR vs. GLIM	GEE vs. GLIM
Intercept	-1.842	0.0311	0.051	0.361	0.310	0.027	0.178	0.151
SEX	-0.126	0.0297	0.003	0.008	0.005	0.026	0.064	0.039
AGE1	-0.155	0.1700	-0.009	-0.027	-0.017	0.061	0.188	0.127
AGE2	-0.531	0.1305	-0.068	-0.040	0.028	0.136	0.079	0.058
AGE3	-0.830	0.0740	-0.117	-0.060	0.057	0.152	0.075	0.076
AGE4	-0.969	0.0504	-0.119	-0.052	0.067	0.131	0.055	0.076
AGE5	-0.912	0.0353	-0.124	-0.063	0.061	0.146	0.071	0.075
AGE6	-0.848	0.0296	-0.116	-0.059	0.057	0.147	0.072	0.075
AGE7	-0.767	0.0269	-0.112	-0.063	0.049	0.157	0.086	0.071
AGE8	-0.596	0.0256	-0.085	-0.050	0.034	0.153	0.088	0.065
AGE9	-0.474	0.0252	-0.083	-0.061	0.022	0.191	0.137	0.055
AGE10	-0.320	0.0246	-0.069	-0.061	0.008	0.241	0.208	0.033
AGE11	-0.215	0.0243	-0.059	-0.058	0.001	0.319	0.310	0.009
AGE12	-0.078	0.0243	-0.034	-0.039	-0.005	0.560	0.681	0.133
AGE13	-0.000	0.0256	-0.020	-0.029	-0.008	2.000	2.000	0.341
AGE15	0.385	0.3071	0.020	0.026	0.006	0.054	0.071	0.016
AGE16	0.447	0.2144	-0.001	-0.025	-0.024	0.003	0.054	0.051
AGE17	0.259	0.1370	-0.009	-0.035	-0.026	0.034	0.126	0.092
AGE18	0.279	0.0867	-0.001	-0.013	-0.012	0.004	0.046	0.042
AGE19	0.312	0.0573	-0.009	-0.034	-0.025	0.028	0.103	0.075
AGE20	0.310	0.0462	0.006	-0.017	-0.024	0.021	0.053	0.075
AGE21	0.248	0.0407	0.012	-0.007	-0.018	0.048	0.027	0.075
AGE22	0.154	0.0379	-0.013	-0.024	-0.011	0.079	0.146	0.066
AGE23	0.154	0.0365	-0.006	-0.018	-0.012	0.037	0.110	0.074
AGE24	0.091	0.0351	-0.010	-0.015	-0.005	0.107	0.154	0.048
AGE25	0.089	0.0342	-0.008	-0.012	-0.004	0.087	0.132	0.044
AGE26	0.054	0.0339	-0.019	-0.022	-0.003	0.300	0.339	0.040
AGE27	-0.041	0.0357	-0.020	-0.015	0.006	0.659	0.439	0.238
C1	-1.312	0.0506	-0.182	-0.053	0.129	0.149	0.041	0.108
C2	-0.948	0.0281	-0.128	-0.043	0.084	0.144	0.047	0.098
C3	-0.117	0.0272	-0.023	-0.013	0.010	0.218	0.114	0.104
C4	0.215	0.0336	-0.037	-0.055	-0.018	0.158	0.227	0.070
C5	-1.314	0.0267	-0.125	-0.011	0.115	0.100	0.008	0.092
C6	-0.639	0.0253	-0.065	-0.014	0.052	0.107	0.021	0.086
C7	0.167	0.0285	0.004	-0.003	-0.006	0.021	0.017	0.038
C8	0.565	0.0435	0.020	-0.001	-0.021	0.036	0.002	0.038
C9	-1.031	0.0417	-0.060	0.024	0.085	0.060	0.023	0.083
C10	-0.453	0.0290	-0.057	-0.023	0.034	0.135	0.052	0.083
C11	0.280	0.0307	-0.031	-0.064	-0.033	0.105	0.206	0.101
C12	0.858	0.0340	-0.017	-0.098	-0.081	0.019	0.108	0.089
C13	0.655	0.0323	0.027	-0.022	-0.049	0.042	0.033	0.075
C14	0.491	0.0304	-0.005	-0.054	-0.049	0.010	0.104	0.094
C15	0.935	0.0257	0.020	-0.063	-0.083	0.022	0.065	0.087
C16	1.061	0.0437	0.015	-0.080	-0.096	0.015	0.073	0.088
C17	-0.946	0.0265	-0.000	0.071	0.071	0.000	0.073	0.073
C18	-0.294	0.0254	-0.013	0.012	0.025	0.045	0.041	0.086
C19	0.217	0.0302	0.023	0.010	-0.013	0.113	0.047	0.066

C20	0.654	0.0664	0.024	-0.010	-0.034	0.038	0.014	0.052
C21	-0.062	0.0235	-0.046	-0.044	0.002	1.207	1.115	0.139

Table 3.6.2b. Average Models Estimates and Relative Standard Errors from Simple Logistic Model Fit, Differences and Relative Differences in Model Estimates from a Specific Pair of Modeling Approaches for the In Patient Quality Indicator IQI-25 (IQI 25 – Bilateral Cardiac Catheterization Rate) For Each Parameter in the Model (Time-Trend)

Time-Trend								
Parameter	**Simple Logistic Regression**		**Difference between model estimates from a pair of modeling approaches**			**Relative Difference between model estimates from a pair of modeling approaches**		
	Estimate	**Std. Error**	**SLR vs. GEE**	**SLR vs. GLIM**	**GEE vs. GLIM**	**SLR vs. GEE**	**SLR vs. GLIM**	**GEE vs. GLIM**
Intercept	-1.763	0.0314	0.053	0.355	0.302	0.029	0.183	0.154
SEX	-0.126	0.0297	0.004	0.010	0.006	0.033	0.075	0.042
AGE1	-0.144	0.1698	-0.005	-0.025	-0.020	0.034	0.189	0.156
AGE2	-0.521	0.1303	-0.061	-0.033	0.028	0.125	0.066	0.059
AGE3	-0.824	0.0740	-0.111	-0.054	0.057	0.145	0.068	0.077
AGE4	-0.969	0.0504	-0.116	-0.049	0.067	0.127	0.052	0.075
AGE5	-0.909	0.0353	-0.120	-0.058	0.062	0.141	0.066	0.075
AGE6	-0.844	0.0296	-0.112	-0.054	0.057	0.142	0.067	0.075
AGE7	-0.766	0.0269	-0.109	-0.060	0.049	0.153	0.081	0.072
AGE8	-0.594	0.0256	-0.081	-0.046	0.035	0.146	0.081	0.065
AGE9	-0.472	0.0252	-0.079	-0.057	0.023	0.183	0.127	0.056
AGE10	-0.319	0.0246	-0.067	-0.058	0.009	0.233	0.199	0.034
AGE11	-0.215	0.0243	-0.058	-0.056	0.002	0.313	0.301	0.011
AGE12	-0.078	0.0243	-0.033	-0.038	-0.005	0.536	0.643	0.117
AGE13	0.000	0.0256	-0.019	-0.027	-0.008	1.907	1.934	0.344
AGE15	0.400	0.3070	0.027	0.043	0.015	0.071	0.114	0.042
AGE16	0.448	0.2143	-0.002	-0.023	-0.022	0.004	0.050	0.047
AGE17	0.262	0.1367	-0.011	-0.037	-0.026	0.041	0.133	0.092
AGE18	0.281	0.0867	-0.002	-0.014	-0.012	0.007	0.048	0.041
AGE19	0.315	0.0572	-0.008	-0.034	-0.025	0.026	0.101	0.076
AGE20	0.310	0.0462	0.006	-0.018	-0.023	0.018	0.056	0.074
AGE21	0.248	0.0407	0.012	-0.007	-0.019	0.048	0.027	0.075
AGE22	0.156	0.0379	-0.013	-0.025	-0.012	0.080	0.151	0.071
AGE23	0.156	0.0365	-0.006	-0.019	-0.013	0.037	0.113	0.076
AGE24	0.092	0.0351	-0.011	-0.017	-0.005	0.116	0.167	0.051
AGE25	0.088	0.0342	-0.008	-0.013	-0.005	0.091	0.137	0.047
AGE26	0.054	0.0339	-0.020	-0.023	-0.004	0.309	0.356	0.048
AGE27	-0.041	0.0357	-0.021	-0.016	0.005	0.691	0.487	0.224
C1	-1.325	0.0506	-0.190	-0.060	0.130	0.154	0.046	0.108
C2	-0.956	0.0281	-0.132	-0.048	0.084	0.148	0.051	0.097
C3	-0.123	0.0272	-0.026	-0.016	0.010	0.237	0.139	0.099
C4	0.212	0.0336	-0.039	-0.058	-0.019	0.169	0.239	0.071
C5	-1.320	0.0267	-0.129	-0.015	0.114	0.103	0.011	0.092
C6	-0.641	0.0253	-0.067	-0.016	0.051	0.110	0.025	0.085
C7	0.165	0.0285	0.003	-0.003	-0.006	0.017	0.019	0.036
C8	0.560	0.0435	0.018	-0.003	-0.021	0.032	0.005	0.038
C9	-1.038	0.0416	-0.065	0.019	0.084	0.065	0.018	0.083
C10	-0.460	0.0290	-0.062	-0.028	0.034	0.144	0.063	0.082
C11	0.278	0.0307	-0.033	-0.065	-0.033	0.110	0.210	0.100
C12	0.856	0.0341	-0.018	-0.098	-0.081	0.020	0.109	0.088
C13	0.647	0.0324	0.024	-0.026	-0.050	0.037	0.040	0.077
C14	0.488	0.0304	-0.007	-0.056	-0.049	0.014	0.108	0.094
C15	0.933	0.0257	0.019	-0.063	-0.082	0.021	0.065	0.086
C16	1.059	0.0437	0.015	-0.080	-0.095	0.014	0.072	0.087

C17	-0.955	0.0265	-0.006	0.066	0.071	0.006	0.067	0.072
C18	-0.299	0.0254	-0.015	0.010	0.025	0.052	0.032	0.084
C19	0.215	0.0303	0.023	0.010	-0.013	0.114	0.050	0.065
C20	0.650	0.0665	0.022	-0.011	-0.033	0.035	0.016	0.051
C21	-0.064	0.0235	-0.048	-0.046	0.002	1.200	1.119	0.122
Year	-0.076	0.0042	-0.001	0.004	0.005	0.013	0.047	0.060

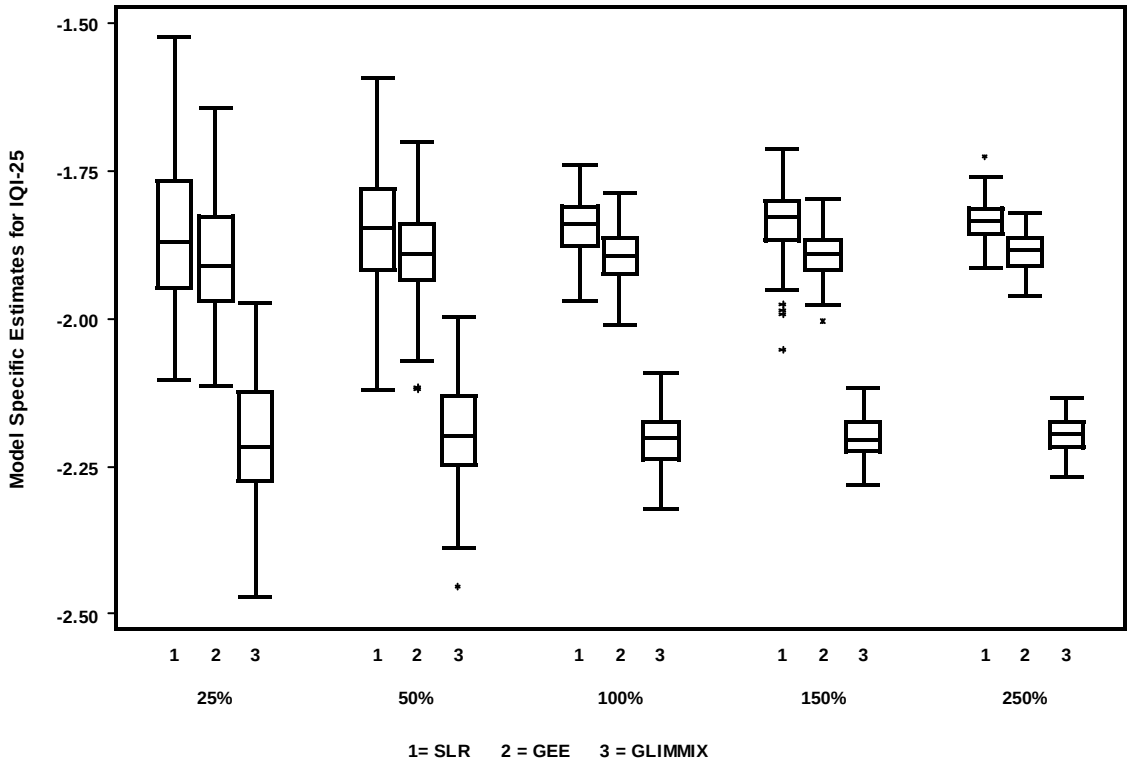


Figure 3.6.2a: Box Plots of the Model Specific Estimates of the Intercepts Obtained from fitting Cross-Sectional models to 100 Boot Strap Sampling with Target Sampling at 25%, 50%, 100%, 150%, and 250% of the Observed NIS Sample Data for IQI-25.

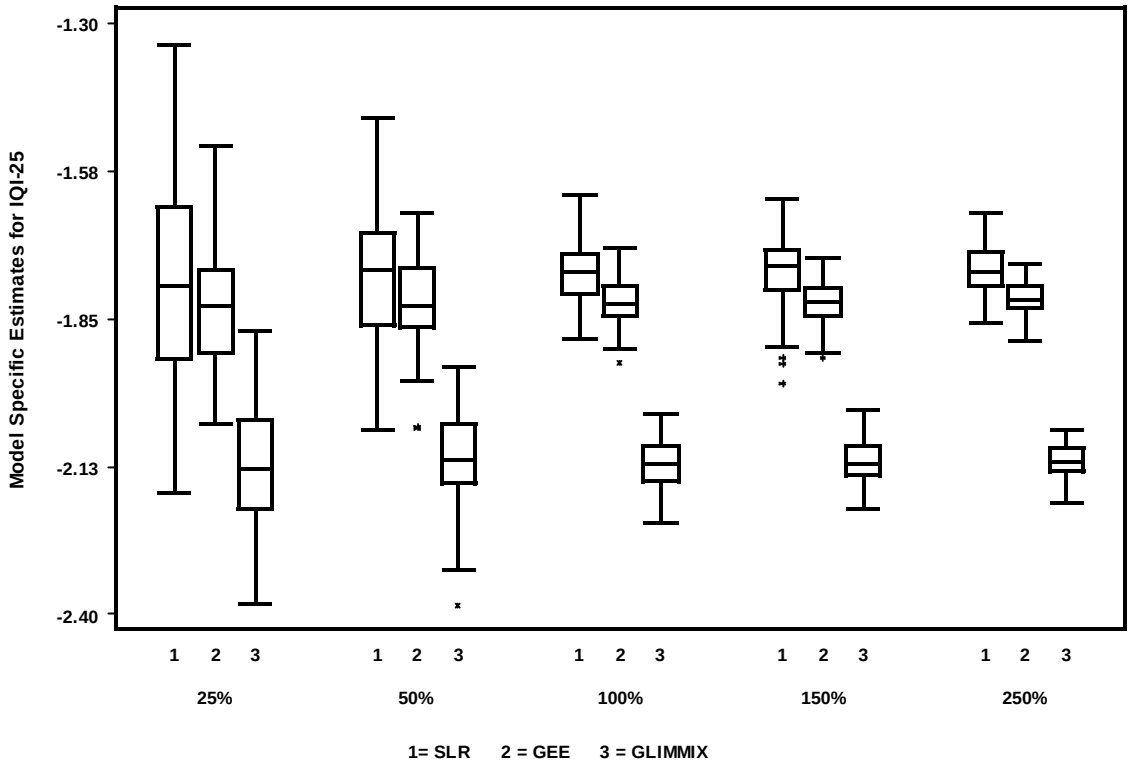


Figure 3.6.2b: Box Plots of the Model Specific Estimates of the Intercepts Obtained from fitting Time-Trend to 100 Boot Strap Sampling with Target Sampling at 25%, 50%, 100%, 150%, and 250% of the Observed NIS Sample Data for IQI-25.

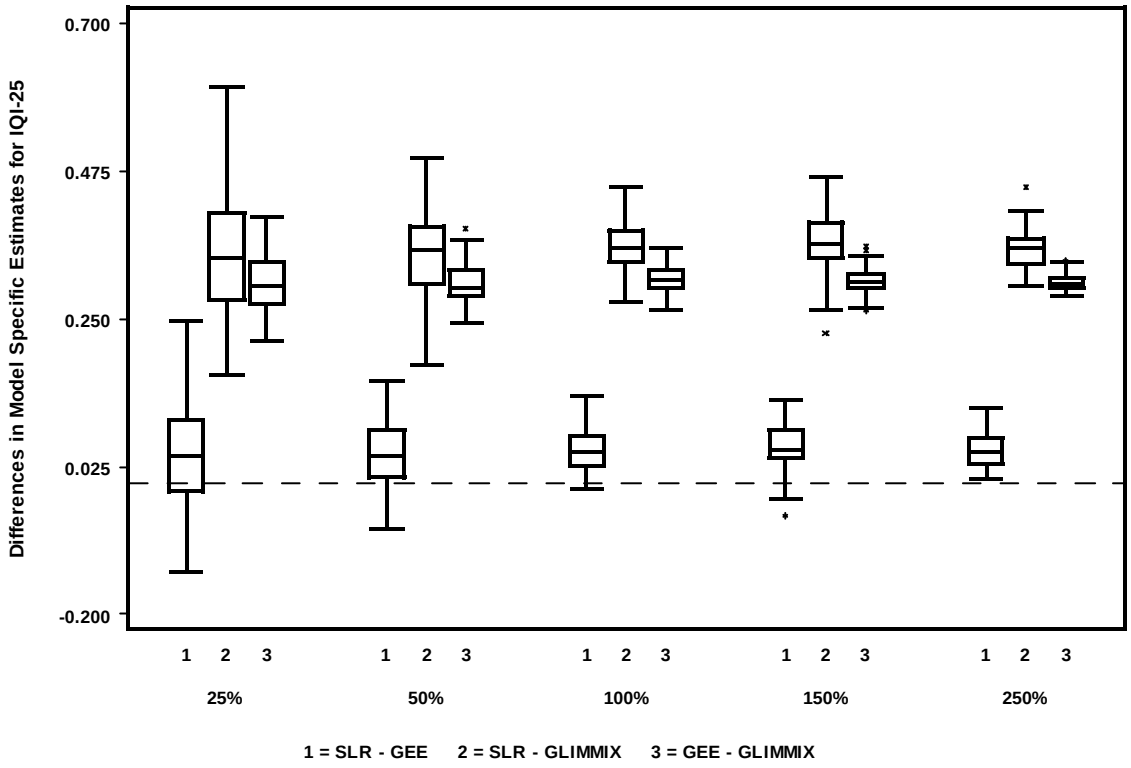


Figure 3.6.2c: Box Plots of the Differences in Model Specific Estimates of the Intercepts from a Specific Pair of Cross-Sectional Models Fitted to 100 Boot Strap Sampling Runs with Target Sampling at 25%, 50%, 100%, 150%, and 250% of the Observed NIS Sample Data for IQI-25.

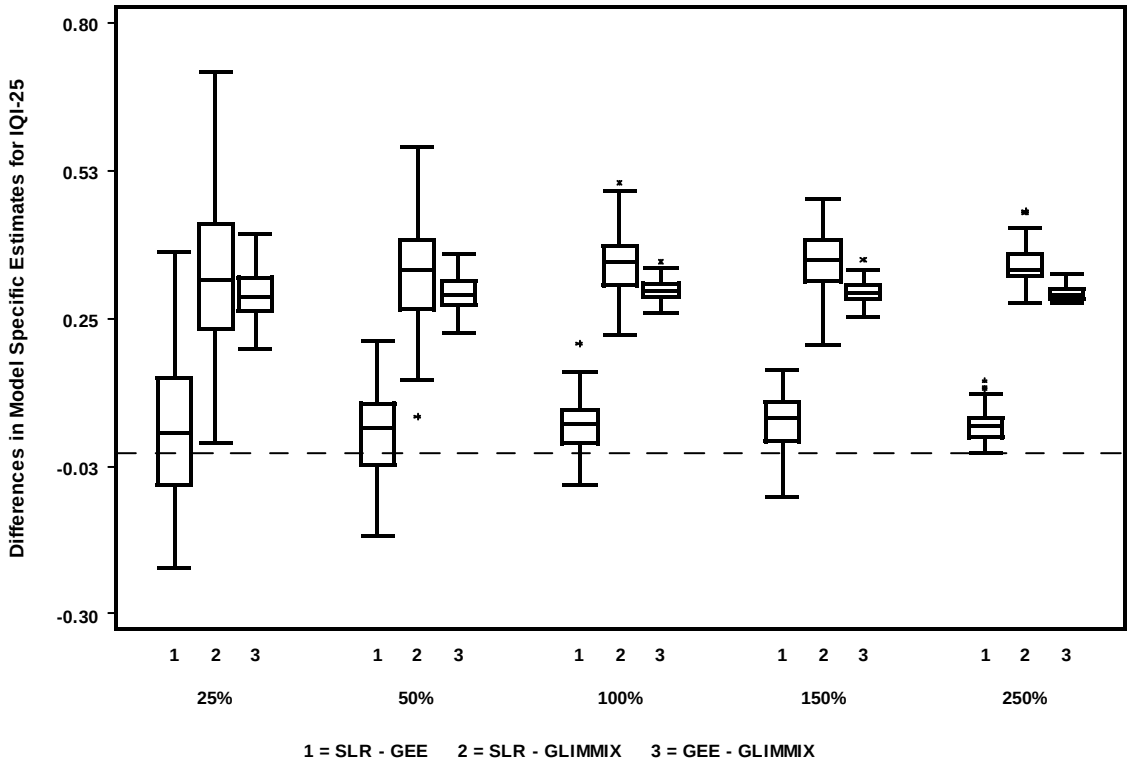


Figure 3.6.2d: Box Plots of the Differences in Model Specific Estimates of the Intercepts from a Specific Pair of Time-Trend Models Fitted to 100 Boot Strap Sampling Runs with Target Sampling at 25%, 50%, 100%, 150%, and 250% of the Observed NIS Sample Data for IQI-25.

Table 3.6.2c Wald Test Statistics and (P-Value) Comparing Models fit to IQI-25 (IQI 25 – Bilateral Cardiac Catheterization Rate)

Test	Percent of Observed Sample	P-values					
		Cross-Sectional Models			Time-Trend Models		
		SLR vs. GEE	SLR vs. GLIM	GEE vs. GLIM	SLR vs. GEE	SLR vs. GLIM	GEE vs. GLIM
Global Wald Test	25%	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
	50%	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
	100%	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
	150%	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
	250%	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
T-Test for Intercept Difference	25%	< 0.0001	< 0.0001	< 0.0001	0.0033	< 0.0001	< 0.0001
	50%	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
	100%	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
	150%	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
	250%	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001

4 Conclusions

The results of this investigation demonstrated that there can be subtle, yet statistically significant differences in the parameter estimates yielded from different modeling approaches for fitting covariate adjusted logistic regression models for the Quality Indicator Project. As anticipated, accounting for positive correlation among patients within the same hospital using the GEE and GLIMMIX approaches resulted in meaningful differences in standard errors associated with the parameter estimates compared to the standard errors yielded from a simple logistic regression model. However, based on analyses applied to the NIS sample data, we also observed statistically significant differences in the vector of parameter estimates in 2 of the five QI's selected for this investigation. This result was not anticipated, as large sample theory suggests that all three methods should converge to the same true parameter values.

Careful inspection of these differences showed that some of these differences, while statistically significant, were very subtle. This is not surprising based on the very large sample size associated with the NIS, which has the effect of magnifying even the smallest of observed differences. For 4 of the 5 QI's investigated, changes in the modeling approach did not create a meaningful difference in the expected rates (relative to the National mean response). However, in many cases, the differences in provider-level estimates of expected and risk-adjusted rates between the modeling approaches were significantly different than zero based on the random samples of 50 providers. This is suggestive of a subtle, yet statistically significant bias.

The results of the bootstrap sampling analyses also demonstrated (and perhaps amplified) the fact that there are differences between the three statistical methods for fitting the covariate adjusted logistic regression models. The bootstrap results showed that variability among parameter estimates shrunk as the sample sizes increased, as expected. However, there were small, yet systematic differences in the parameter estimates between the methods that were found to be statistically significant.

It is also not outside the realm of possibility that the observed differences in the parameter estimates are attributable to choices of convergence criteria and/or estimation method between the SLR, GEE, and GLIMMIX when implemented in SAS. These models are compute intense and difficult to fit – and further work to investigate the cause of differences between model fitting methods in application to the AHRQ QI project is recommended.

From a certain perspective, the simple logistic regression model can be thought of a special case of the GEE approach (in which the correlation coefficient is constrained to a value of zero) and of the GLIMMIX approach (in which the variance component(s) associated with the random effects are also constrained to a value of zero). Of course, in practice, we allow these parameters to be unconstrained with the data suggesting whether the correlation coefficient (GEE) or variance components (GLIMMIX) depart from zero. In this case, we clearly can recommend either the GEE or GLIMMIX approach as being

superior to the simple logistic regression modeling approach because they will properly account for the potential effects of positive correlation among patients treated within the same provider.

The choice between the GEE and GLIMMIX approaches is more difficult, with both approaches offering distinct advantages. The GEE approach is more likely to converge in all cases, using a marginal model with a method of moments estimator to handle the within-provider correlation. The GLIMMIX approach may require more careful modeling of the data, but offers more flexibility in using the variance components to characterize the distribution of quality among the population of observed providers. Through appropriate interpretation of the variance components from the GLIMMIX modeling approach, the AHRQ methodology may be augmented to allow individual providers to assess where they might be located within the National distribution of providers (rather than a simple comparison of whether they are above or below the National mean response).

Finally, the simple adjustment for a linear trend over time resulted in a highly significant negative slope for all QI's investigated. Inclusion of covariates in future QI models to adjust for changes in quality of care over time is recommended as a method for improving the model prediction in application. Addition of time-trend estimation within the QI's may also provide AHRQ and other users with valuable insight into how rates are changing over time both Nationally, and among individual providers via interpretation of the random effect variance components from the GLIMMIX model.

References

AHRQ (2005) Guide to Prevention Quality Indicators. Technical Report published by the U.S. Department of Health and Human Services Agency for Healthcare Research and Quality, accessed on 8/28/06 at http://www.qualityindicators.ahrq.gov

McCullagh, P. & Nelder, J.A. (1989) Generalized Linear Models. 2nd edition. London: Chapman and Hall.

Liang, K.Y. & Zeger, S.L. (1986) Longitudinal data analysis using generalized linear models. Biometrika. 73:13-22.

Royall, R.M. (1986) Model robust inference using maximum likelihood estimators. International Statistical Review. 54:221-226.

Zeger, S.L. & Liang, K.Y. (1986) Longitudinal data analysis for discrete and continuous outcomes. Biometrics. 42:121-130.

Fitzmaurice, G.M., Laird, N.M., and Ware, J.H. (2004) Applied longitudinal analysis. John Wiley and Sons.

Risk Adjustment and Hierarchical Modeling Workgroup

Workgroup Members

Dan R. Berlowitz, Bedford Veterans Affairs Medical Center
Cheryl L. Damberg, Pacific Business Group on Health
R. Adams Dudley, Institute for Health Policy Studies, UCSF
Marc Nathan Elliott, RAND
Byron J. Gajewski, University of Kansas Medical Center
Andrew L. Kosseff, Medical Director of System Clinical Improvement, SSM Health Care
John Muldoon, National Association of Children's Hospitals and Related Institutions
Sharon-Lise Teresa Normand, Department of Health Care Policy Harvard Medical School
Richard J. Snow, Doctors Hospital, OhioHealth

Liaison Members

Simon P. Cohn, National Committee on Vital and Health Statistics (Kaiser Permanente)
Donald A. Goldmann, Institute for Healthcare Improvement
Andrew D. Hackbarth, Institute for Healthcare Improvement
Lein Han, Centers for Medicare & Medicaid Services
Amy Rosen, Bedford Veterans Affairs Medical Center
Stephen Schmaltz, Joint Commission on Accreditation of Healthcare Organizations

Technical Advisors

Rich Averill, 3M
Robert Baskin, AHRQ
Norbert Goldfield, 3M
Bob Houchens, Medstat
Eugene A. Kroch, Institute for Healthcare Improvement Technical Advisor (Carescience)

AHRQ QI Support Team

Agency for Healthcare Research and Quality:
Mamatha Pancholi
Marybeth Farquhar

Battelle Memorial Institute:
Warren Strauss
Jyothi Nagaraja
Jeffrey Geppert
Theresa Schaaf

APPENDIX –A

Generalized Estimating Equations

In many studies, we are faced with a problem where the responses Y_i are not independent of each other ($Cov[Y_i,Y_j] \neq 0$ when $i \neq j$). The responses from studies with correlated data can often be organized into clusters, where observations from within a cluster are dependent, and observations from two different clusters are independent:

Y_{ij} is the j^{th} response from the i^{th} cluster: $Cov[Y_{ij},Y_{i'j'}]= 0$ when $i \neq i'$
$Cov[Y_{ij},Y_{ij'}] \neq 0$ when $j \neq j'$

In the context of the AHRQ Quality Indicators project, the providers (hospitals) serve as clusters. There are usually two objectives for the analysis of clustered data:

1) Describing the response variable Y_{ij} as a function of explanatory variables (X_{ij}),and
2) Measuring the within-cluster dependence.

When Y_{ij} is continuous and follows a normal distribution, there is a well developed set of statistical methodology for meeting the above two objectives. This methodology usually assumes that the residuals from within each cluster are jointly normal, so that each cluster is distributed $MVN(X_i\beta, \Sigma_i)$. Thus, when faced with normally distributed dependent responses, the assumption of a multivariate normal distribution allows us to model clustered data with our usual maximum likelihood solutions.

When the response variable does not follow a normal distribution, we are often left without a multivariate generalization which allows us to meet the two objectives for the analysis of clustered data through use of a maximum likelihood solution. For example, there are no multivariate extensions of the binomial distribution that provides a likelihood function for clustered data.

The theory of Generalized Estimating Equations (GEE) provides a statistical methodology for analyzing clustered data under the conceptual framework of Generalized Linear Models. GEE was developed in 1986 by Kung-Yee Liang and Scott Zeger, and is an estimating procedure which makes use of Quasi-Likelihood theory under a marginal model.

When the regression analysis for the mean is of primary interest, the β coefficients can be estimated by solving the following estimating equation:

$$U_1(\beta,\alpha) = \sum_{i=1}^{K} \left(\frac{\partial \mu_i}{\partial \beta}\right)' cov^{-1}(Y_i;\beta,\alpha)(Y_i - \mu_i(\beta)) = 0$$

where $\mu_i(\beta) = E[Y_i]$, the marginal expectation of Y_i

Note that U_1 (the GEE) has exactly the same form as the score equation from a simple logistic regression model, with the exception that:

1) Y_i is now an $n_i H1$ vector which comprises the n_i observations from the i^{th} cluster
2) The covariance matrix, $cov(Y_i)$, for Y_i depends not only on β, but on α which characterizes the within cluster dependence.

The additional complication of the parameter α can be alleviated by iterating until convergence between solving $U_1(\beta, \alpha(\beta)) = 0$ and updating $\alpha(\beta)$, an estimate of α. Thus, the GEE approach is simply to choose parameter values β so that the expected $\mu_i(\beta)$ is as close to the observed Y_i as possible, weighting each cluster of data inversely to its variance matrix $cov(Y_i;\beta,\alpha)$ which is a function of the within-cluster dependence.

The marginal GEE approach has some theoretical and practical advantages:

1) No joint distribution assumption for $Y_i = (Y_{i1},...,Y_{ini})$ is required to use the method. The GEE approach utilizes a method of moments estimator for α, the within-cluster dependence parameter.

2) β, the solution of $U_1(\beta, \alpha(\beta)) = 0$, has high efficiency compared to the maximum likelihood estimate of β in many cases studied.

3) Liang and Zeger have proposed the use of both a model-based and a robust variance of β. The model-based variance of β is more efficient – but is sensitive to misspecification of the model for within-cluster dependence. The robust variance is less efficient, but provides valid inferences for β even when the model for dependence is misspecified.

   Specifically, suppose the investigators mistakenly assume that the observations from the same cluster are independent from each other. The 95% confidence interval for each regression coefficient β_j, j=1,..,p, based upon $\beta_j \forall 1.96 (V_\beta)^2$ remains valid in large sample situations. Thus, investigators are protected against misspecification of the within-cluster dependence structure. This is especially appealing when the data set is comprised of a large number of small clusters.

When the within-cluster dependence is of primary interest, this marginal GEE approach has an important limitation – in that β and α are estimated as if they are independent of each other. Consequently, very little information from β is used when estimating α.

The marginal model for correlated binary outcomes (such as those from the AHRQ QI Project) can be thought of as a simple extension to a simple logistic regression model that directly incorporates the within-cluster correlation among patient responses from within the same hospital. To estimate the regression parameters in a marginal model, we make assumptions about the marginal distribution of the response variable (e.g. assumptions about the mean, it's dependence on the explanatory variables, the variance of each Y_{ij},

and the covariance among responses from within the same hospital). The cross-sectional model (Model (1)) and time-trend model (Model (3)) can be fit using the generalized estimating equations approach using SAS Proc Genmod, through the introduction of a repeated statement that accounts for the within-provider clustering.