AHRQ Present on Admission (POA) – Technical Overview

Jeffrey Geppert, EdM, JD, Battelle Memorial Institute

Warren Strauss, ScM, Battelle Memorial Institute

John Bott, MSSW, MBA, Contractor, Agency for Healthcare Research and Quality

May14, 2010

2:00 to 3:30 pm ET

Toll Free: 1-877-939-8827; Passcode: AHRQ POA

AHRQ QI 2010 webinars to date

- **January 12 and 14**
  - AHRQ QIs, Version 4.1 – Overview
- **January 25 and 27**
  - AHRQ QIs, Version 4.1 – Additional Detail
- **May 12**
  - AHRQ QIs use of Present on Admission – User Overview
- **May 14**
  - AHRQ QIs use of Present on Admission – Technical Overview

Agenda

- POA Overview (5 minutes)
  - Approach
- POA Model Steps (25 minutes)
  - Statistical Notation
  - Goal
  - Assumptions
  - Bayesian Approach
  - Data Imputation
  - Markov chain Monte Carlo (MCMC) Analysis
- Software Tool (5 minutes)
- Example (5 minutes)
- Discussion (30-45 minutes)

POA Overview: Approach

- Two sets of algorithms needed to incorporate POA information
  - 1. Develop response variables and comorbidity factor covariates in the **presence** of POA data
    - Less measurement error thereby more accurate and based on fewer assumptions
  - 2. Develop response variables and comorbidity factor covariates in the **absence** of POA data
    - Use observed POA data to estimate probability of POA for response and comorbidity factors for patients that do not have POA data
    - Provide hospital with risk-adjusted rate that would be "most likely" had they collected POA data
- Observed and estimated data are used to develop the final AHRQ QI models

POA Model Steps: Statistical Notation

- Y_{ij} = PSI Indicator for the j^{th} patient in the i^{th} hospital
  - Y_{ij}=1 if the patient experiences the adverse health effect, 0 otherwise
- P_{ij} =Indicator of whether the adverse health effect (represented by Y_{ij}) is present on admission - determined from the POA data.
  - Note that P_{jj} will equal 0, by definition, if Y_{ij} =0, but that P_{ij} could equal either 0 or 1 when Y_{ij} =1. P_{ij} is not observed on everyone.
- Z_{ij} = Vector of explanatory variables associated with the j^{th} patient in the i^{th} hospital, based on administrative records with no POA data.
  - Z_{ij} is observed for everyone.
- X_{ij} =Vector of improved explanatory variables associated with the j^{th} patient in the i^{th} hospital, based on administrative records with POA data.
  - X_{ij} is not observed on everyone.

POA Model Steps: Goal

- Our goal is to predict:

$$\Pi_{ij} = \Pr(Y_{ij} = 1 \mid P_{ij} = 0, X_{ij})$$

POA Model Steps: Assumptions

- Assume: $Logit(\pi_{ij}) = X_{ij}\beta_Y + \delta_{Y,i}$

- Subcomponent of the model is prediction of : $r_{ij} = \Pr[P_{ij} = 1 | X_{ij}]$

  - Assume: $Logit(r_{ij}) = X_{ij}\beta_P + \delta_{P,i}$

POA Model Steps: Assumptions (cont.)

- Account for the anticipated within-hospital correlation among Y_{ij} responses, using a Generalized Estimating Equations (GEE) Approach
  - A random effects approach was considered, but was discarded because multiple observed hospitals with no cases were compromising the random effect estimates

Likelihood Equations

- If POA data are available (and hence x_{ij} and P_{ij} are observed), we maximize the following likelihood, where r_{ij} is the probability that P=1, given the observable characteristics of X.

$$L_{ij} = \left(\pi_{ij}^{1-P_{ij}}\right)^{Y_{ij}} \left(1-\pi_{ij}^{1-P_{ij}}\right)^{1-Y_{ij}} r_{ij}^{P_{ij}}(1-r_{ij})^{1-P_{ij}}[X_{ij}, Z_{ij}] \quad (i,j) \in \Omega,$$

- When x_{ij} and/or P_{ij} is not observed, we need to integrate/sum over the missing data P and X. Information about both of these may be obtained in the variables Z that are generally observed.

$$L_{ij} = \int_{X_{ij}} \sum_{P_{ij}} \left(\pi_{ij}^{1-P_{ij}}\right)^{Y_{ij}} \left(1-\pi_{ij}^{1-P_{ij}}\right)^{1-Y_{ij}} r_{ij}^{P_{ij}}(1-r_{ij})^{1-P_{ij}} g(X_{ij}, Z_{ij}) \quad (i,j) \notin \Omega,$$

Bayesian Approach

- Because X_{ij} can be >100, the integral equation is unfeasible
- To avoid calculating the integral
  - Use the following approach:

$$[\theta \mid W] \propto [W \mid \theta][\theta]$$

- If direct calculation of the likelihood is unfeasible can use MCMC sampling

Data Estimation

- Combined use of Bayesian approach and other sampling techniques is convenient for missing data
- If $\{\theta^{(j)}, W'^{(j)}\}) j = 1,...n$ is a random sample from $[\theta, W' \mid W] \propto [W, W', \theta]^{\xi}$ then $\{\theta^{(j)}\}) j = 1,...n$ is a random sample from $[\theta \mid W] \propto [W, \theta] = \int_{W'} [W, W', \theta]$
- Allows sampling of augmented posterior distribution [W, W', θ] rather than integration over missing data

Model Fitting Approach using MCMC Overview

- Multiple pre-processing steps prior to fitting
  - Ensures data are formatted and sorted as anticipated
  - Eliminates columns of Z (and X) that are linearly dependent with each other
  - Allows for multiple P variables (i.e., P_1, P_2, P_3) – where $P=Max(P_k)$

- MCMC Approach
  1. Establish X|Z using a series of 2x2 tables, and Establish P|X using a logistic regression modeling approach
  2. Impute values of X where missing using X|Z, and impute values of P where missing using P|X – creating an MCMC simulated analysis dataset
  3. Establish Y|X,P=0 by fitting the logistic regression model Y|X for the subset of the MCMC simulated analysis dataset in which P=0.
     - Repeat steps 2-3 many times until parameter estimates reach convergence

- Analysis module fits the models two ways – using a Naïve simple logistic regression modeling approach, and using a GEE approach that accounts for within-hospital correlation

Model Fitting Approach using MCMC Overview (cont.)

- Begin with

$$[Y, P, X, Z, P', X' \mid X' \mid \beta_Y, \delta_Y, \beta_P, \delta_P, \beta_Z] = [X \mid X'][P \mid P'] \times [Y, P', X'Z]\beta_Y, \delta_Y, \beta_P, \delta_P, \beta_Z$$

- P' and X' indicate the "true process, such that Y = Y' and Z = Z' always, while for P and X, we set

$$[X \mid X'][P \mid P'] = \prod_{(i,j)\in\Omega} \delta(X_{ij} - X'_{ij}) \prod_{(i,j)\in\Omega} \delta(P_{ij} - P'_{ij})$$

Model Fitting Approach using MCMC (cont.)

- For the process model:

$$\left[Y, P', X', Z \mid \beta_Y, \delta_Y, \beta_P, \delta_P, \beta_Z\right] = \prod_{ij}\left[Y_{ij} \mid P'_{ij}, X'_{ij}, \beta_Y, \delta_{Y,i}\right]$$

$$\times \prod_{ij}\left[P' \mid X'_{ij}, \beta_P, \delta_{P,i}\right]$$

$$\times \prod_{ij}\left[X' \mid Z_{ij}, \beta_{x,ij}\right]$$

$$\times \left[Z\right]$$

Model Fitting Approach using MCMC (cont.)

- with $[Y_{ij}|P'_{ij},\beta_Y,\delta_{Y,i}] = (\pi_{ij}^{1-P'ij})^{Yij}(1-\pi_{ij}^{1-P'ij})^{1-Yij}$

$$[P'_{ij} \mid X'_{ij}, \beta_P \delta_{P,i}] = r_{ij}^{P'ij}(1-r_{ij})^{1-P'ij}$$

$$[X'_{ij} \mid Z'_{ij}, \beta_{X,ij}] = s_{ij}^{Zij}(1-s_{ij})^{1-Zij}$$

- Where π_{ij} and r_{ij} and $Logit(s_{ij}) = \beta_{X,ij0} + Z_{ij}\,\beta_{x,ij1}$

Model Fitting Approach using MCMC (cont.)

- Large number of parameters in the augmented likelihood and high percentage of missing P and X data, MCMC sampling may be unstable
  - Not representative of posterior parameter distribution
  - Need simplified model
    - Use logistic regression on subset of sample that have no missing data to consider them fixed during MCMC simulation
    - Gibbs (instead of Metropolis-Hasting)

Model Fitting Approach using MCMC (cont.)

- Consider normal asymptotic expansion for fixed effects of logistic regression

$$\prod_{ij}(\pi_{ij}^{1-P'ij})(1-\pi_{ij}^{1-P'ij})^{1-Yij} \approx N(\beta_Y, \delta_Y);(\hat{\beta}_Y\hat{\delta}_Y), \hat{\Sigma})$$

- Run logistic regression on left hand side of equation above to create a set of parameters for normal function of right hand side of equation

Model Fitting Approach using MCMC (cont.)

- To account for random hospital effects generalized estimated equations (GEE) theory is used to account for within-hospital correlation:

$$\prod_{ij}(\pi_{ij}^{1-P'ij})^{Yij}(1-\pi_{ij}^{1-P'ij})^{1-Yij} \approx N(\beta_Y; \hat{\beta}_{GEE}, \hat{\Sigma}_{GEE})$$

Model Fitting Approach using MCMC (cont.)

- In sum,
  - Use component-wise Metropolis-Hasting sampler, draw "true process" variables P' and X' according to augmented likelihood (use both data-model equations and process-model equations for estimation)
  - Use component-wise Gibbs sampler, draw fixed effect β_Y using a GEE normal approximation

Model Fitting Approach using MCMC (cont.)

- Linear Dependence
  - Use singular value decomposition (SVD) to decompose the matrices: X^TX and Z^TZ using the kernel of matrix M = kernel of M^TM.
- Separation
  - MLE approach produces infinite estimates for certain fixed effects
  - Use regularization term: ridge regression
  - "Flat" normal prior distribution:

$$[\beta_Y] = N(\beta_Y; 0, \mathrm{I}/\lambda)$$

Model Fitting Approach using MCMC (cont.)

- Estimate doesn't affect β_Y but stabilized the solution
- Improves instability due to residual collinearity in the data (i.e., not removed by SVD)

Software Development

- Two Modules have been developed that implement the POA-Adjusted Quality Indicator Models
  - Analysis Module for fitting National data from the HCUP
    - Provides parameter estimates and associated standard errors from Naïve and GEE-based MCMC models:

Y|Z	Similar to previously developed AHRQ Models
P|X	Based on data where X is Observed
Y|X, P=0	Based on data where X & P are Observed
Y|X, P=0 (MCMC)	Based on Imputed data across entire dataset

  - Prediction Module for applying Model Results to patient records from a select Hospital (or group of Hospitals)
    - Uses consistent MCMC approach to impute values of P and X (where missing) prior to applying parameter estimates – averaging the predicted values of Y over many simulations

Software Development (cont.)

- C++ program that implements MCMC simulations (patented by Battelle)
- Reads comma separated file containing the Y, P, X and Z data
- Eliminates zero and linearly dependent columns
- Performs GEE regression analyses on the distributions noted on last slide

Software Development (cont.)

- Once coefficients fitted as appropriate performs standard and GEE analysis through MCMC with data estimation of the distribution [Y|P=0,X]
- With more POA indicators a univariate value is calculated
- GEE regression (not MCMC simulation) "model standard errors" and "empirical standard errors" are calculated

Software Development (cont.)

- After analysis, hospital predictions are calculated
- Software inputs: data filename; number of POA indicators; pathname of folder to store results; result filenames (standard regression analysis, GEE regression analysis, standard prediction results and GEE prediction results); subfolders to store various analytic files; analysis values, parameters and analytic steps; and, name of file to store log

Software Development (cont.)

- Prediction Module
  - The software tool can perform hospital aggregate predictions and individual predictions based on previous analyses

Example

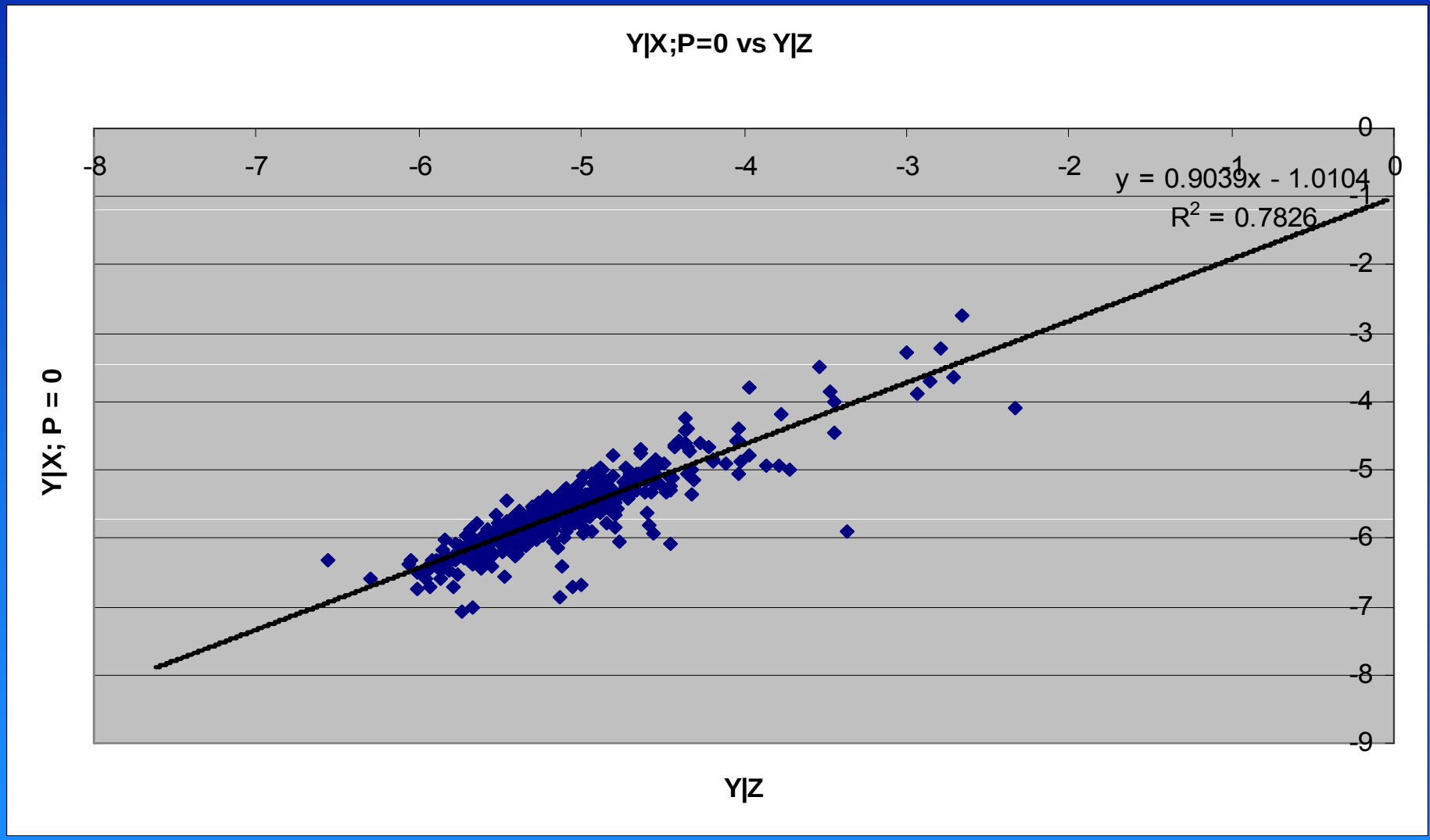
- Postoperative Sepsis

	Discharges without POA Data	Discharges with POA Data		
tpps13/ qpps13 (P)	**Missing**	**0**	**1**	**Total**
0	549,614	248,629	0	798,243
1	8,208	2,312	1,436	11,956
Total	557,822	250,941	1,436	810,199
0	98.53%	98.51%	0.00%	98.51%
1	1.47%	0.92%	0.57%	1.49%
Total	100.00%	99.43%	0.57%	100.00%

Table B2. Number and Percent of Discharges by Flag
Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2007. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/sidoverview.jsp.
Note: tpps13 = inclusion in numerator; qpps13 = inclusion in denominator; (P) = cases flagged in outcome of interest excluded from population at risk because outcome is POA; 0 – does not meet inclusion; 1 = meets inclusion.

Results: Postoperative Sepsis (current approach)

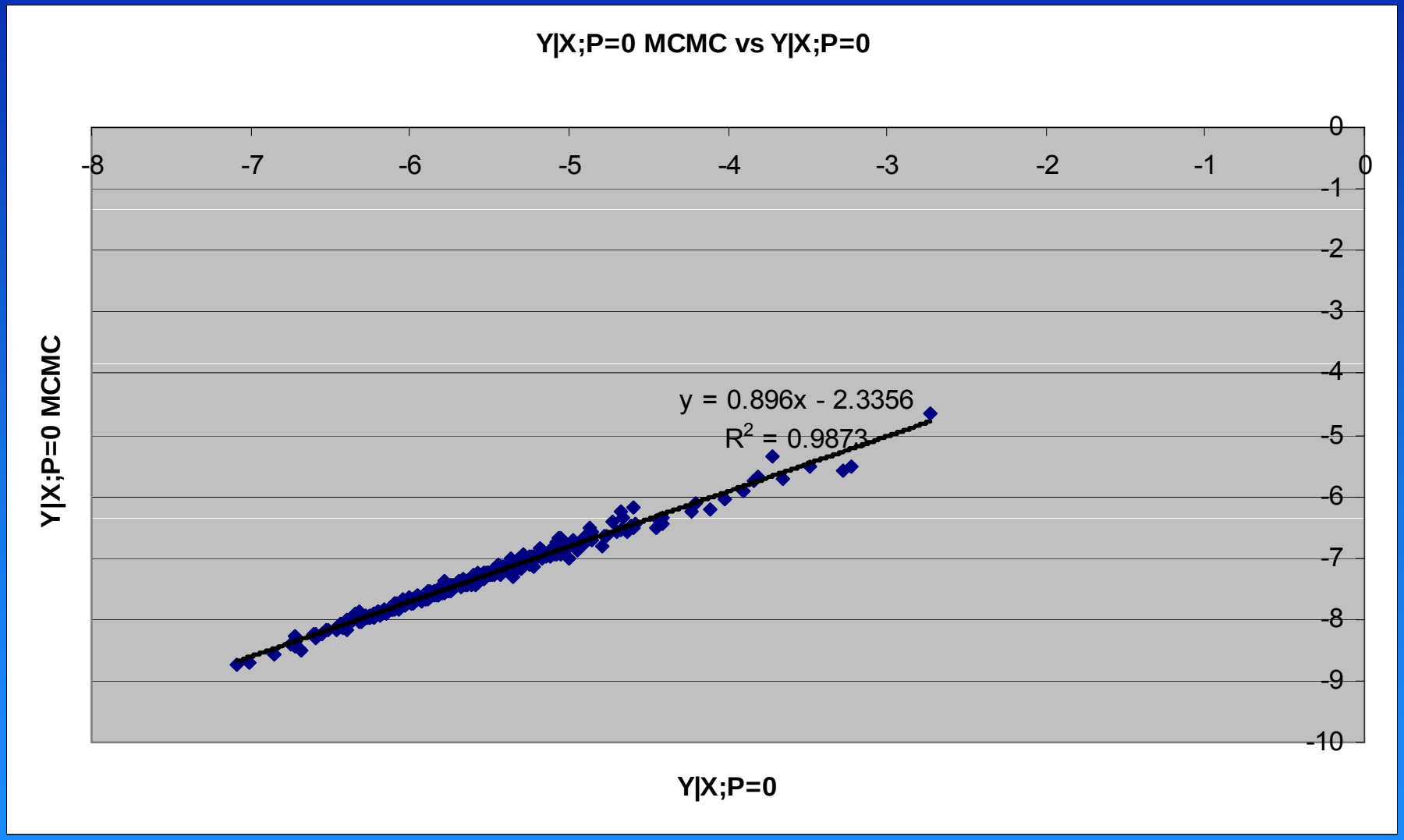
With POA



Without POA

Results: Postoperative Sepsis (Alternative approach)

New Model



With POA

Some Potential Next Steps

- Continue to refine the AHRQ QI numerator, denominator and risk factor definitions
  - Improve the sensitivity and specificity of the indicators
- Incorporate other tools to improve the coding of present on admission
  - Publicly available diagnostics on the accuracy of POA coding

Discussion

- For your consideration:
  - Did this webinar meet your needs?
    - Content? Scope?
  - How will the information presented be useful to you?
  - Is there anything we did not cover or didn't address in enough detail for you?

- Your questions:
  - Questions about what you heard today?
    - If we don't answer your question today, then we will post a response on the AHRQ QI website

For more information…

AHRQ QIs

- Web site: http://qualityindicators.ahrq.gov/
  - AHRQ QI documentation and software are available at the AHRQ QI web site
- Present on Admission White Paper:
  - http://www.qualityindicators.ahrq.gov/downloads/webinars/Using%20Present%20on%20Admission.pdf
- Staff:
  - Mamatha Pancholi Mamatha.Pancholi@ahrq.hhs.gov
  - John Bott John.Bott@ahrq.hhs.gov