A Method for Obtaining Stable Physician/Clinic Performance Estimates

How many patients/charts do I need to adequately represent physician and clinic performance in a trial?

R. L. Brown, Ph.D.

Professor

Schools of Nursing and Medicine

University of Wisconsin-Madison

Presented at 14th Annual Primary Care Research Methods & Statistics Conference, San Antonio, Texas, Dec. 3-5, 1999

Issues Regarding the Sampling Design of Primary Care Trials

- What is the unit of analysis, randomization, and intervention?
  - Single-level trial?
  - Multi-level trial?
    - Patient outcome measures
    - Intervention directed toward physicians
    - Randomization of clinics

Issues Regarding the Sampling Design of Primary Care Trials

- Protocol specific study
  - addiction medicine study concerned with only a few clustered outcomes
  - dependency of outcomes
- General protocol study
  - wide array of preventive outcomes

Issues Regarding the Sampling Design of Primary Care Trials

- Measurement issues
  - ceiling effects on certain protocols (e.g., smoking screening)
- Nested homogeneity in multi-level studies
  - dependency due to simple nesting
  - dependency due to clinic characteristics

Issues Regarding the Sampling Design of Primary Care Trials

- How large should my sampling be?
- How should the population characteristics be brought into the sampling strategy?
- How reliable are these estimates?

Definitions

- Power: The probability of rejecting the null hypothesis when it is false.
- Reliability: The extent to which an experiment, test, or any measuring procedure yields the same results on repeated trials.
- Single-level analysis: Analysis based on a single group (e.g., patients or physicians).
- Multilevel analysis: Analysis of nested groups (e.g., patients, nested under physicians, physicians nested under clinics).

Sampling Strategies

- Single-level assumed probability.
- Single-level unknown/uncertain probability.
- Two-level randomization/intervention of physician with patient as proxy - unknown probability - unconditional model.
- Two-level randomization/intervention of physician with patient as proxy - unknown probability - conditional model.

Sampling Strategies (cont)

- Three-level randomization/intervention of clinic with nested physicians and patient as proxy - unknown probability - unconditional model.
- Three-level randomization/intervention of clinic with nested physicians and patient as proxy - unknown probability - conditional model.
- Multiple membership models.

Single-level trial with known probability

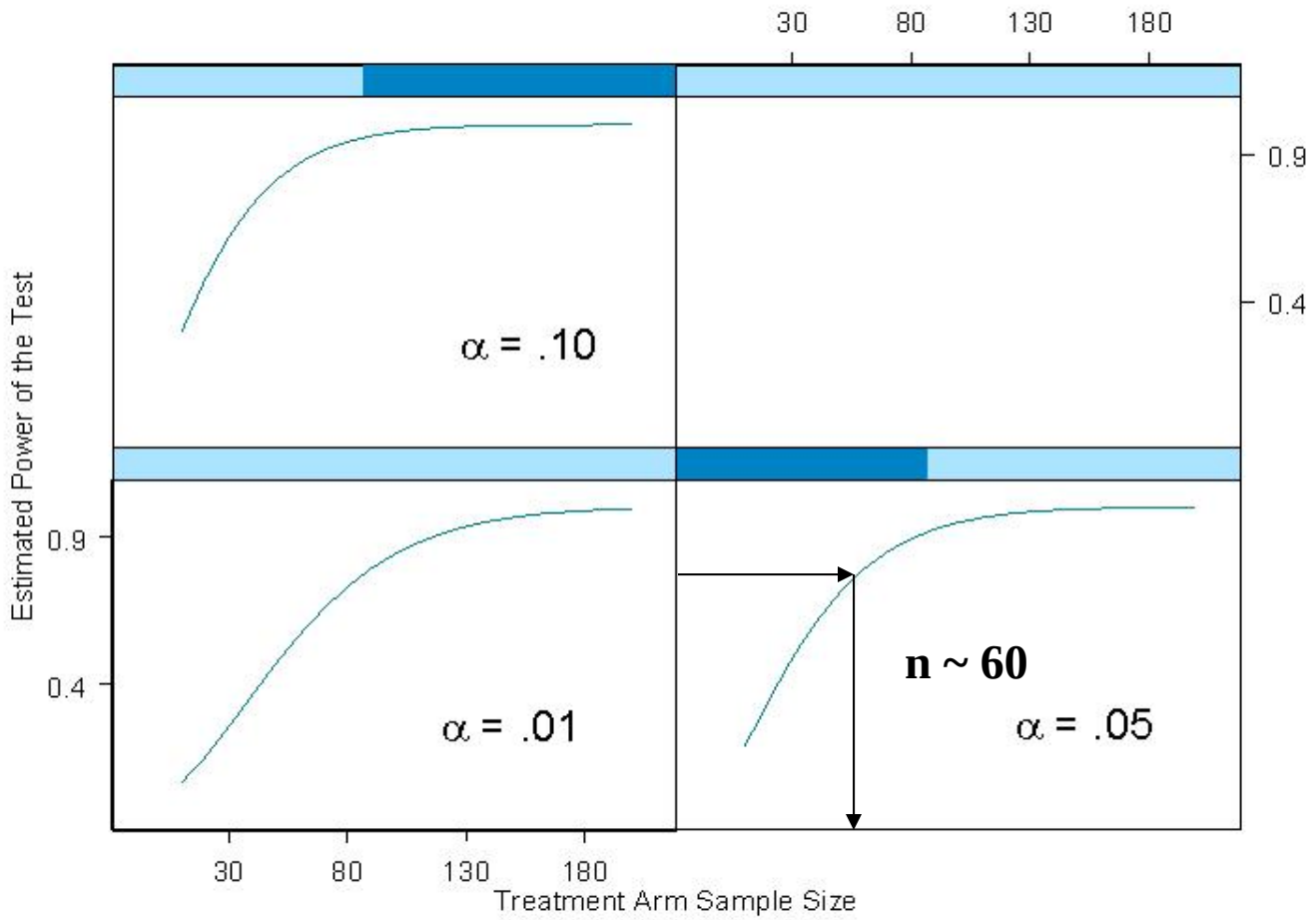
This trial assumes a stable known probability of the population characteristic, so sampling for this trail becomes a function of power only.

If the trial wishes to test:

$$H_o: \text{p1} = \text{p2} \text{ versus } H_A: \text{p1} <> \text{p2}$$

The sampling strategy is fairly straight forward and the sample size may be estimated using a variety of methods (e.g., arcsine transformation approach).

Estimated Power Functions Based on Arcsine Transformation



Single-level trial with unknown/uncertain probability

How does one estimate the probability?

- Previous literature
- Clinic judgement (Guess)
- Pilot study

How reliable is my guess or pilot value?

In typical single-level trials (e.g., patient performance) with uncertain probabilities, one may estimate a confidence in the probability.

Typical approximation approach based on random sampling is given by Tryfos (1996).

- Some initial probability value
- Size of the population
- Some level of precision
- Some level of confidence placed on the approximation.

$$n = \frac{N\pi(1-\pi)}{(N-1)D^2 + \pi(1-\pi)}$$

Where N = size of the population, D = (c/$Z_{\alpha/2}$), with c = precision factor, $Z_{\alpha/2}$ is a known factor depending on 1 - α (confidence level), and π = estimated proportional value.

Precision Factor

The precision factor (c) refers to the estimation range (for example, plus or minus θ percentage points) within which the performance estimate derived from examination of sample cases can be assumed to reflect the quality of the population of cases sampled.

Confidence Level

The factors $Z_{\alpha/2}$ corresponding to commonly used $1 - \alpha$ are:

$1 - \alpha$	$Z_{\alpha/2}$	$1 - \alpha$	$Z_{\alpha/2}$
0.99	2.576	0.80	1.282
0.95	1.960	0.60	0.842
0.90	1.645	0.50	0.674

Based on the Central Limit Theorem

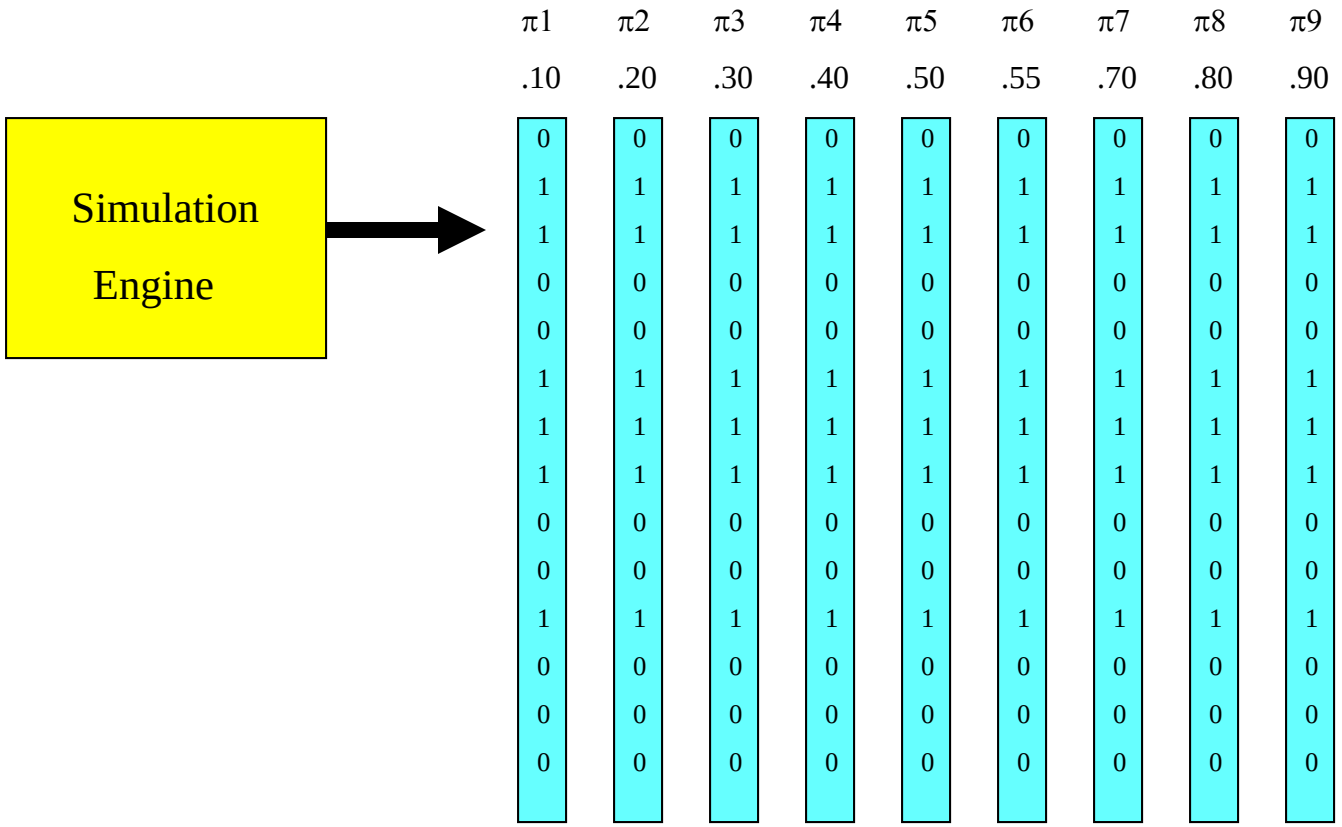
Example

How large a random sample without replacement should be taken of a clinic system with 1,000 patient records so that the estimate of the proportion of patients asked about smoking status (documented on the chart) with an interval of precision of plus or minus 2% with probability of 95%?

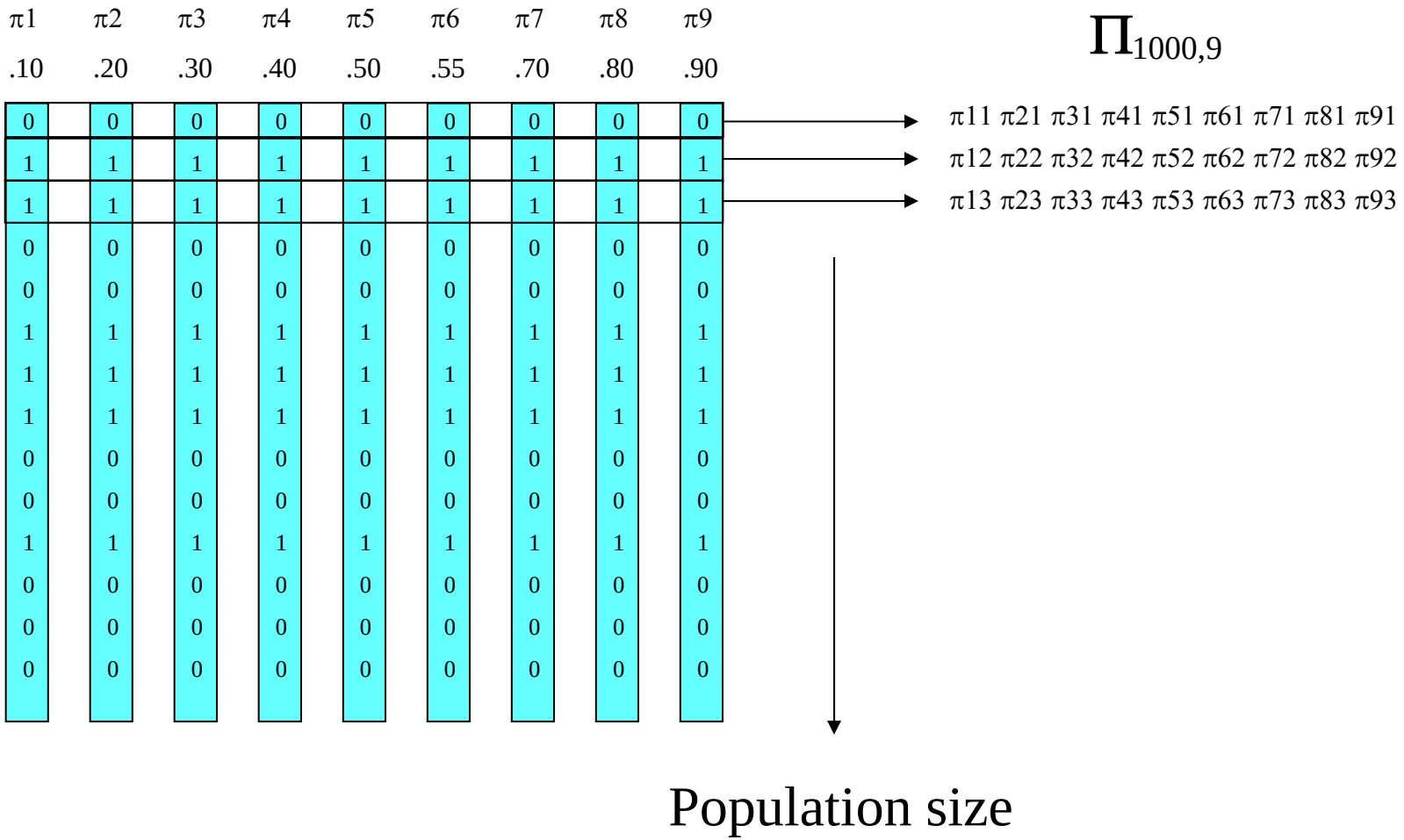
π becomes a critical value in this calculation. If we assume that π = .10, .20, .30, .40, and .50 and base our calculations on a population of 1000 charts, with anticipated precision level of 2% at 95% confidence, we project that we will need to randomly sample the following:

π	$\pi(1 - \pi)$	Projected Number of charts needed
0.1 or 0.9	0.09	463
0.2 or 0.8	0.16	606
0.3 or 0.7	0.21	668
0.4 or 0.6	0.24	697
0.5	0.25	706

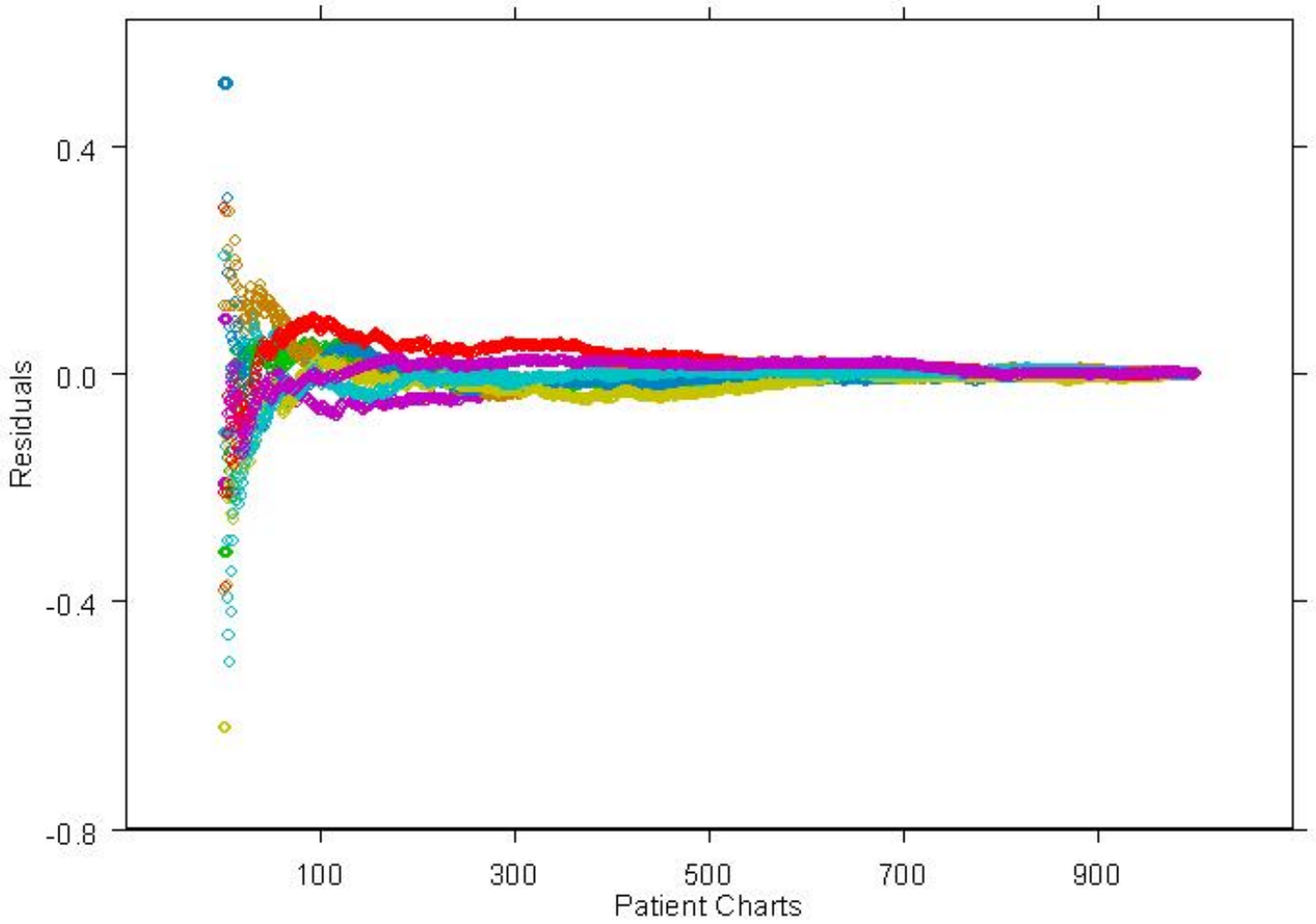
Sequential Computer Simulation

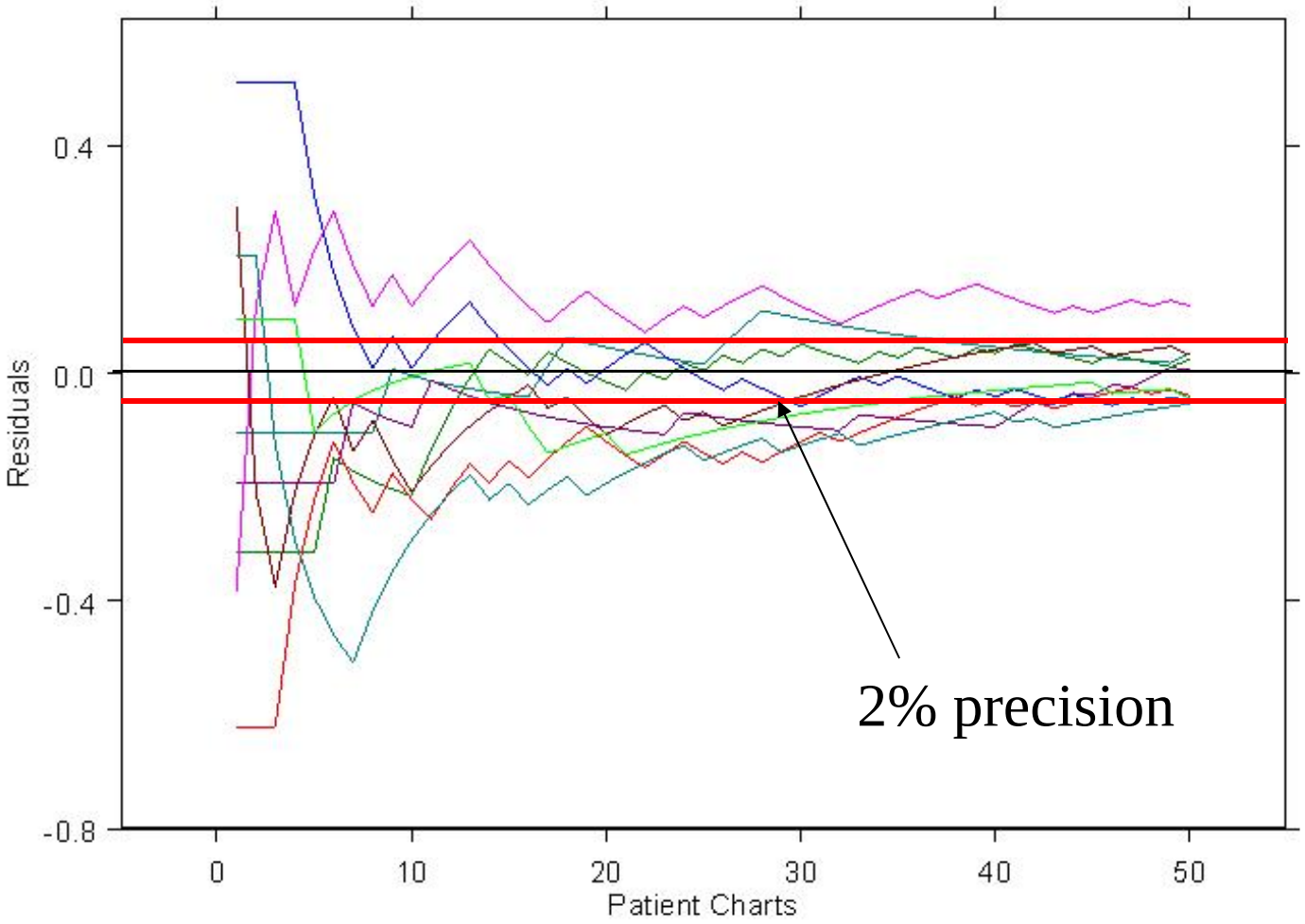


Sequential Calculation Routine



Plot of Estimated Probabilities versus “true” probability



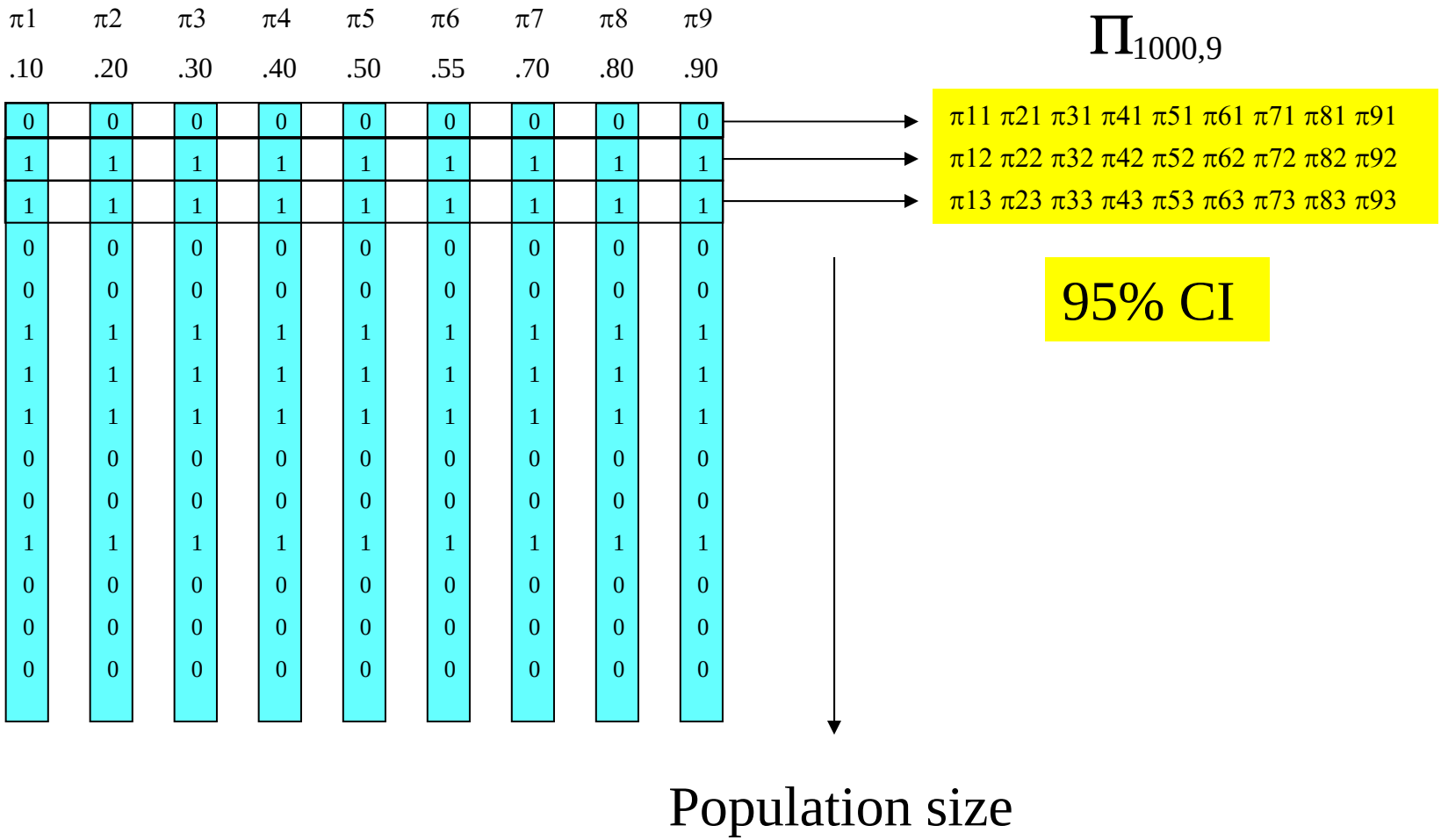

0.4
0.0
-0.4
-0.8
Residuals
0
10
20
30
40
50
Patient Charts
2% precision

Confidence Intervals

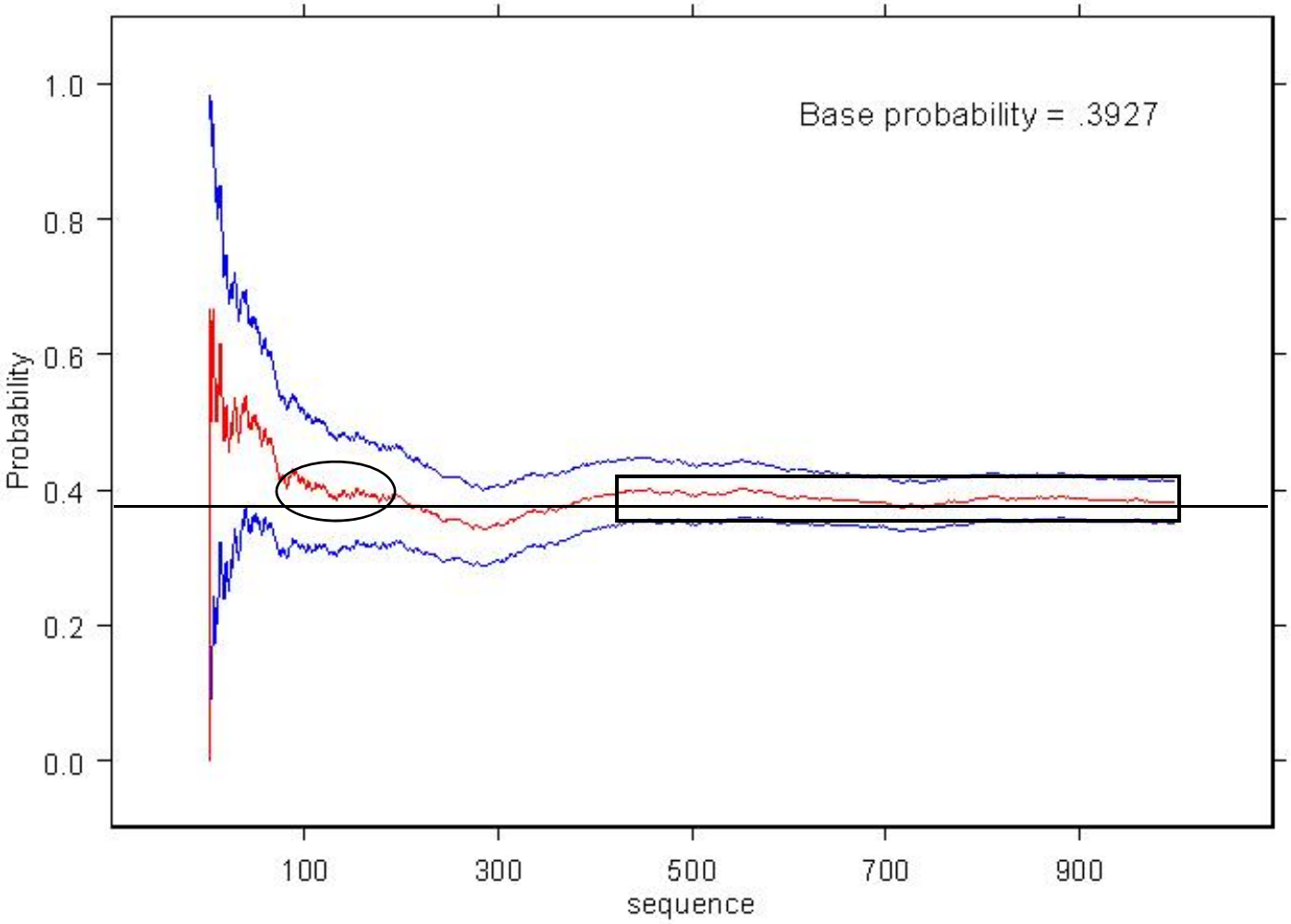
$$p_l = \frac{\left(2np + c^2_{\alpha/2} - 1\right) - c_{\alpha/2}\sqrt{c^2_{\alpha/2} - \left(2 + 1/n\right) + 4p(nq+1)}}{2\left(n + c^2_{\alpha/2}\right)}$$

$$p_u = \frac{\left(2np + c^2_{\alpha/2} + 1\right) + c_{\alpha/2}\sqrt{c^2_{\alpha/2} + \left(2 - 1/n\right) + 4p(nq-1)}}{2\left(n + c^2_{\alpha/2}\right)}$$

Sequential Calculation Routine



95% Confidence Intervals

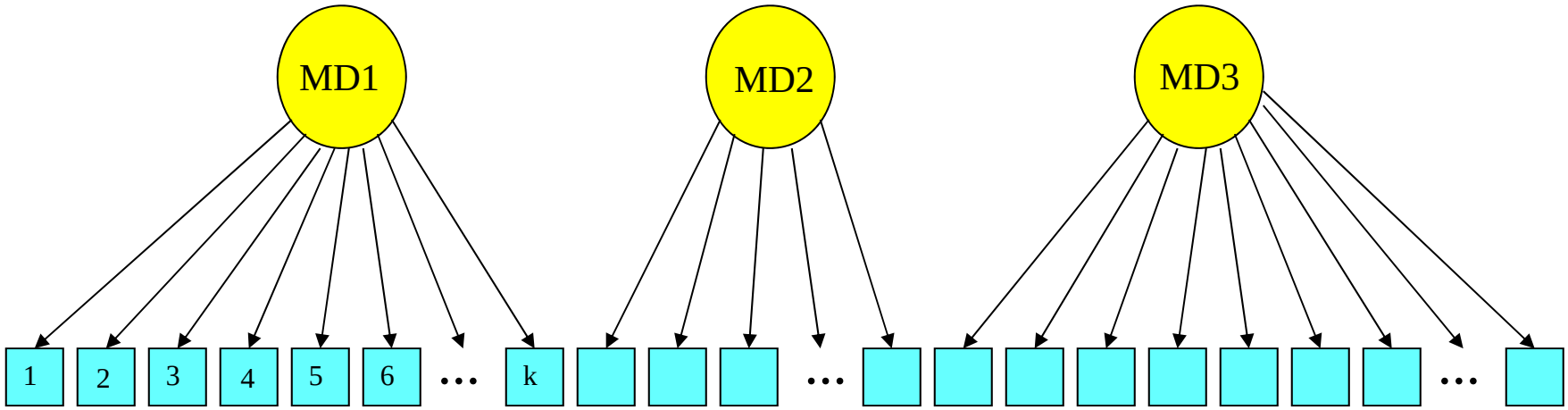


The majority of primary care intervention studies are considered hierarchical in nature

If physician performance is of interest, that performance is usually a function of patient activity, meaning patients seen by that physician are a proxy for physician performance.

We now have a sampling situation where not only an estimate of the probability of an event is important, but that probability may vary across the higher levels (e.g., physicians, clinics, etc.)

Two-Level Structure



Methodology of Variance Components in the General Linear Model

To test the reliability or stability of the scores, one must account for factors, which affect the measurements obtained. This may be the issue of nesting plus a variety of population characteristics.

Variance may arise from multiple factors (e.g., physician characteristics, patient severity, clinic conditions, etc.).

Variance components modeling provides a method for isolating and quantifying multiple sources of variance that impact on measuring the primary care performance.

The goal of this approach is to identify variance from sources other than the physician/clinic and to minimize their contribution to the performance measure.

Central to the VC approach is the idea of fixed and random effects.

To qualify as a *fixed effect*, the levels of an independent variable are selected systematically. In contrast, to qualify as a *random effect*, the levels are selected randomly or unsystematically from a pool of possible levels.

For example, a researcher who want to assess an intervention (*fixed effect*) may choose as a second independent variable clinics within a particular state, but wants to extend conclusions to the population of clinics not participating in the study. Thus clinic becomes a random factor.

What makes a random effect different is that each level of a random effect contributes an amount that is viewed as a sample from a population of normally distributed variables, with mean of 0, and variance of σ^2.

The estimate of σ^2 associated with the *random effect* is known as the variance component because it measures the part of the overall variance contributed by that effect.

The initial step in the VC modeling approach is to identify factors that are potential *random* contributors of variation in the performance score, and the relationships between these factors.

These relationships fall into two categories, “nested” and “crossed” designs. A nested design occurs when measurements on one factor of interest are a subset of some other factor. A crossed design occurs when elements of two or more factors are independent entities but can be measured in all possible combinations of the individual elements.

Methodology of Variance Components in Multilevel Models

Since the majority of primary care research is inherently hierarchical in nature, intervention analyses have been concentrating on hierarchical or multilevel analyses. The sampling strategy must parallel the hierarchical nature.

Two-level randomization/intervention of physician with patient as proxy - unknown probability - unconditional model

What we have proposed has been to sequentially sample in a pilot study the structure of the analysis to help answer questions about sample size and sampling strategy. Since many of our study outcomes are binary (e.g., was a procedure completed or not, or was a history taken), hierarchical logisitic models will provide the framework for the remainder of this session.

A two-level logistic model may be defined as:

$$logit\pi_{ij} \equiv \ln\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = f_{ij} + r_j$$

Where:

π_{ij} is the probability of the i^{th} patient obtaining a procedure from the j^{th} physician.

f_{ij} = fixed factors

r_j = random terms per physician-level

By rewriting the model as below and defining a constant vector of 1's in this model (X_o) with the link function being a binomial we are able to estimate an unconditional two-level model providing two parameter estimates of the overall expectancy or grand probability βo and an estimate of between physician variability σ^2_{u2}

$$logit(\pi_{ij}) = \beta_o X_o + u_j$$

where $u_j \approx N(0, \sigma^2_{u2})$

We then estimate this model in sequential steps beginning with i=1 patients per physician and increment the nested patients by i +1 per physician

To estimate this unconditional two-level logit model I will be using the MLwiN software by Goldstein, et al (1998)

MLwiN: Goldstein, et al.

HLM: Bryk, Raudenbush, Seltzer & Congdon

MLwiN Screen Definition for the two-level unconditional logit model

$$\left.\begin{aligned} \text{outcome}_{patient,\,md} &\sim \text{Binomial}(\text{denom}_{patient,\,md},\ \pi_{patient,\,md}) \\ \text{outcome}_{patient,\,md} &= \pi_{patient,\,md} + e_{1patient,\,md}\text{bcons}^{*} \end{aligned}\right\}$$

$$\text{logit}(\pi_{patient,\,md}) = \beta_{0md}\text{cons}$$

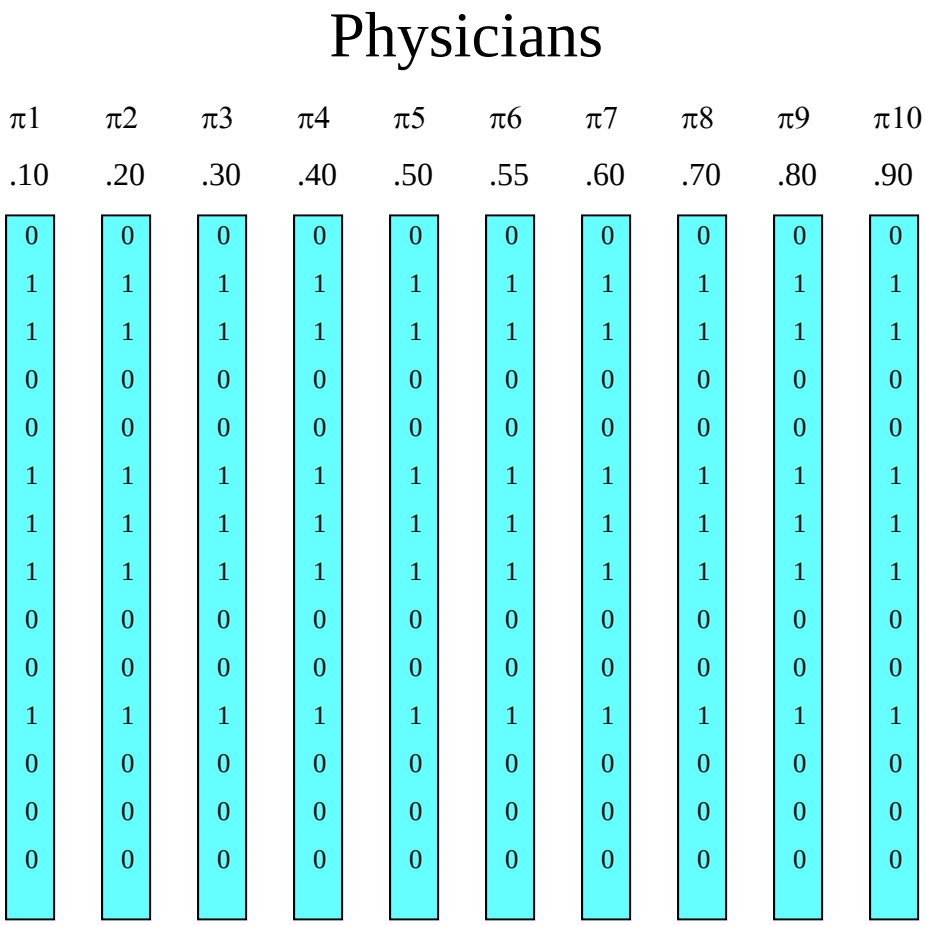
$$\beta_{0md} = 0.055(0.302) + u_{0md}$$

$$\begin{bmatrix} u_{0md} \end{bmatrix} \sim \text{N}(0,\ \Omega_u) \ :\ \Omega_u = \begin{bmatrix} 0.901(0.406) \end{bmatrix}$$

$$\text{bcons}^{*} = \text{bcons}[\pi_{patient,\,md}(1 - \pi_{patient,\,md})/\text{denom}_{patient,\,md}]^{0.5}$$

$$\begin{bmatrix} e_{1patient,\,md} \end{bmatrix} \sim (0,\ \Omega_e) \ :\ \Omega_e = \begin{bmatrix} 1.000(0.000) \end{bmatrix}$$

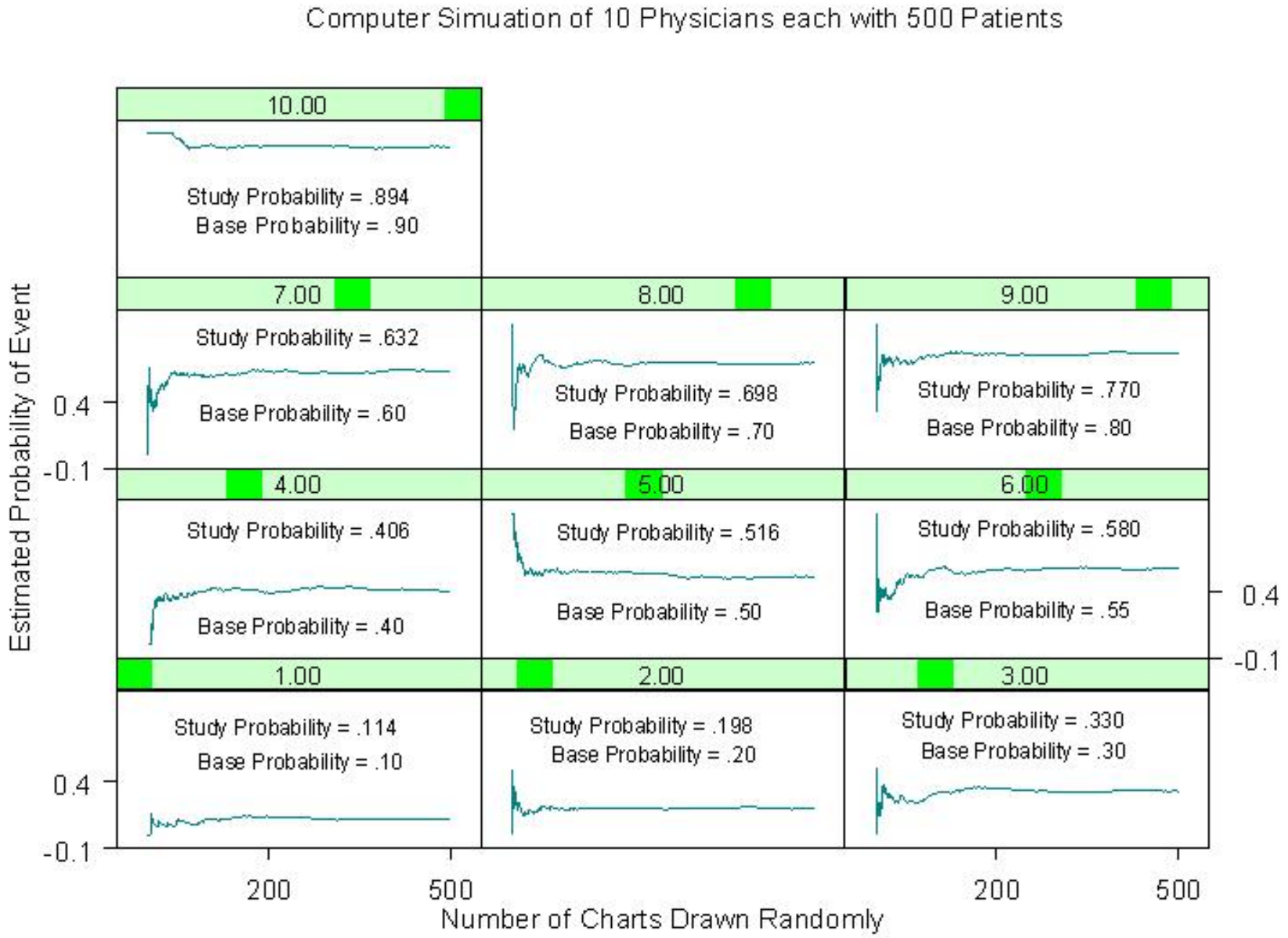
Computer Simulation Study

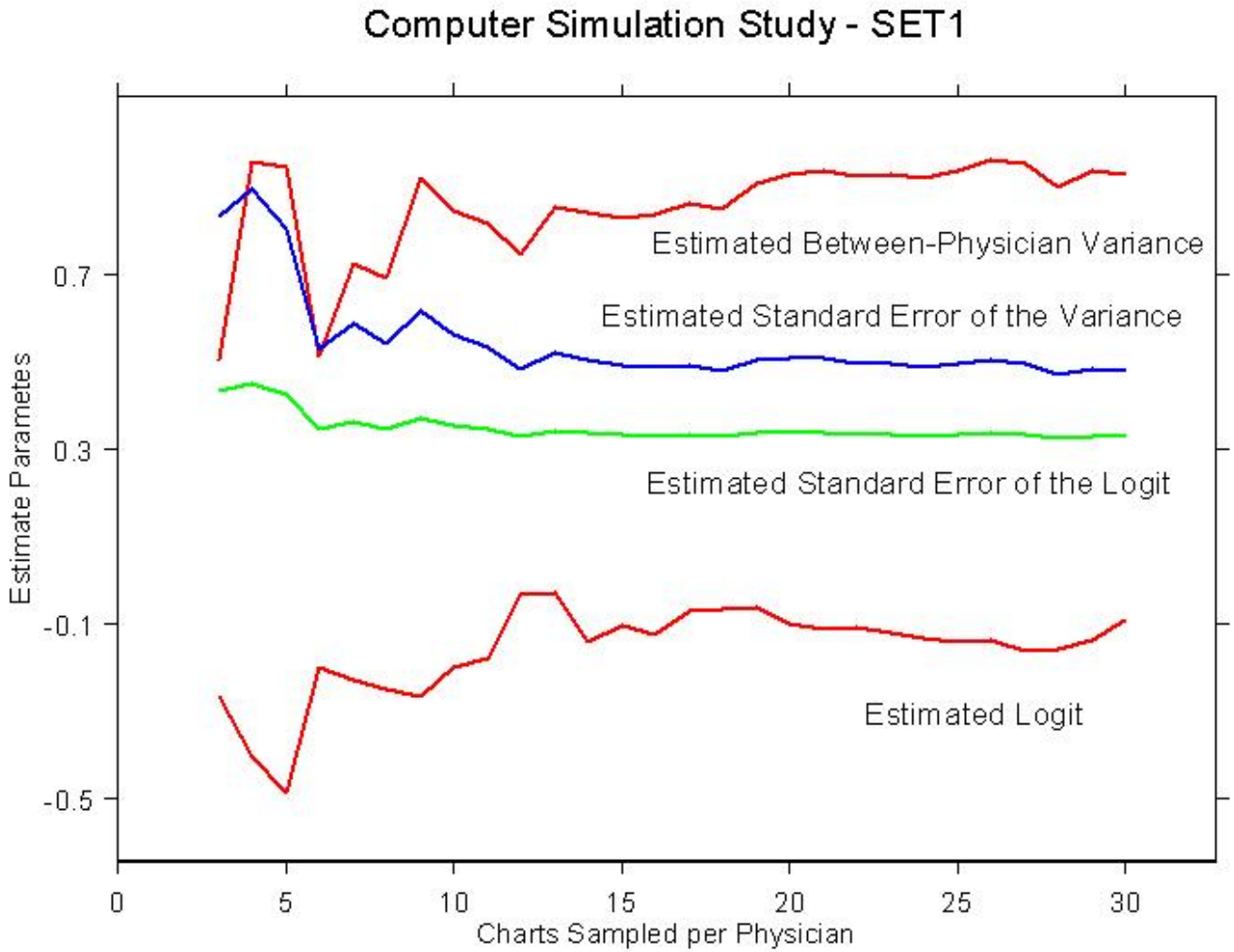


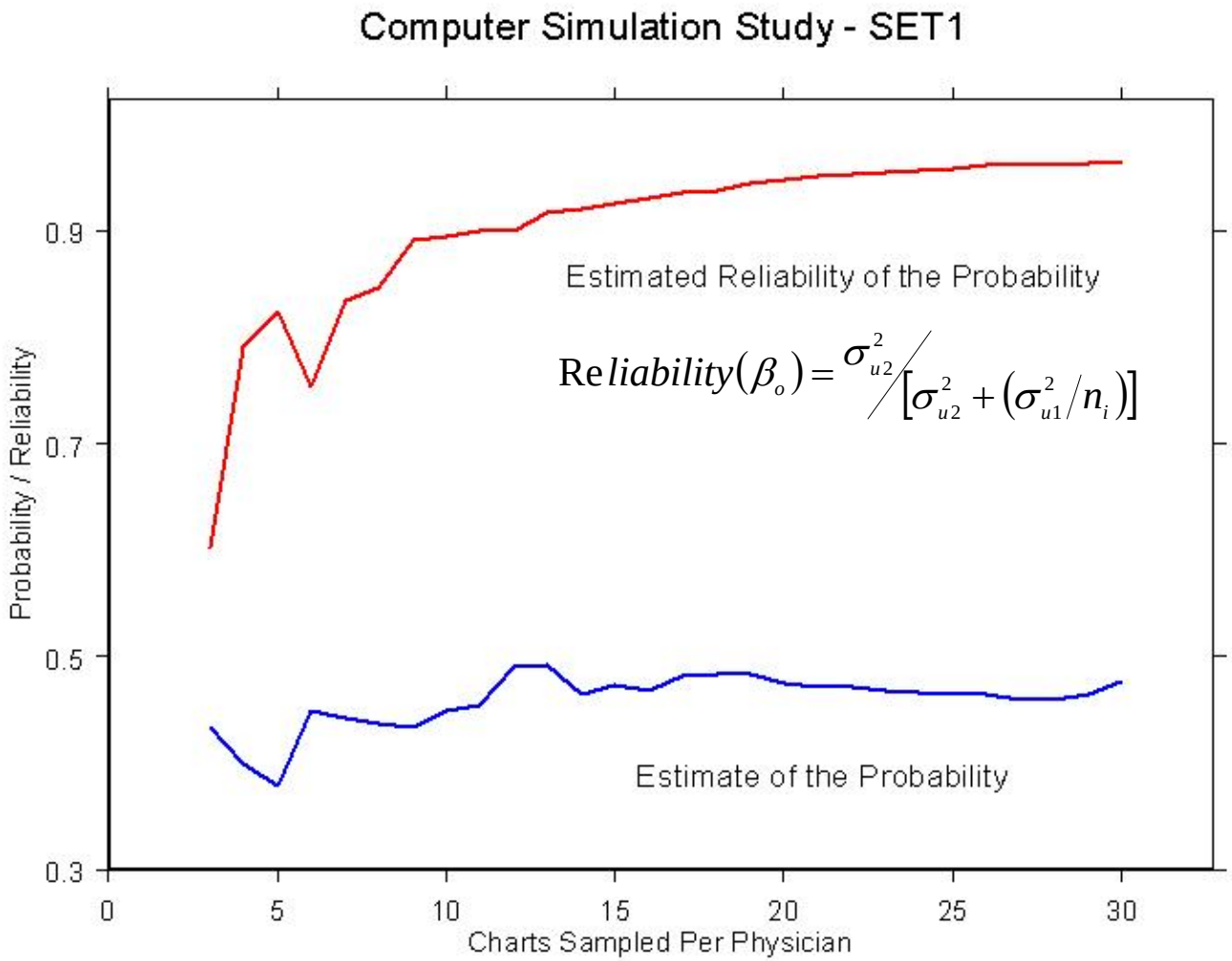
$\Pi_{500,10}$

$\pi_{11}\ \pi_{21}\ \pi_{31}\ \pi_{41}\ \pi_{51}\ \pi_{61}\ \pi_{71}\ \pi_{81}\ \pi_{91}\ \pi_{101}$
$\pi_{12}\ \pi_{22}\ \pi_{32}\ \pi_{42}\ \pi_{52}\ \pi_{62}\ \pi_{72}\ \pi_{82}\ \pi_{92}\ \pi_{102}$
$\pi_{13}\ \pi_{23}\ \pi_{33}\ \pi_{43}\ \pi_{53}\ \pi_{63}\ \pi_{73}\ \pi_{83}\ \pi_{93}\ \pi_{103}$

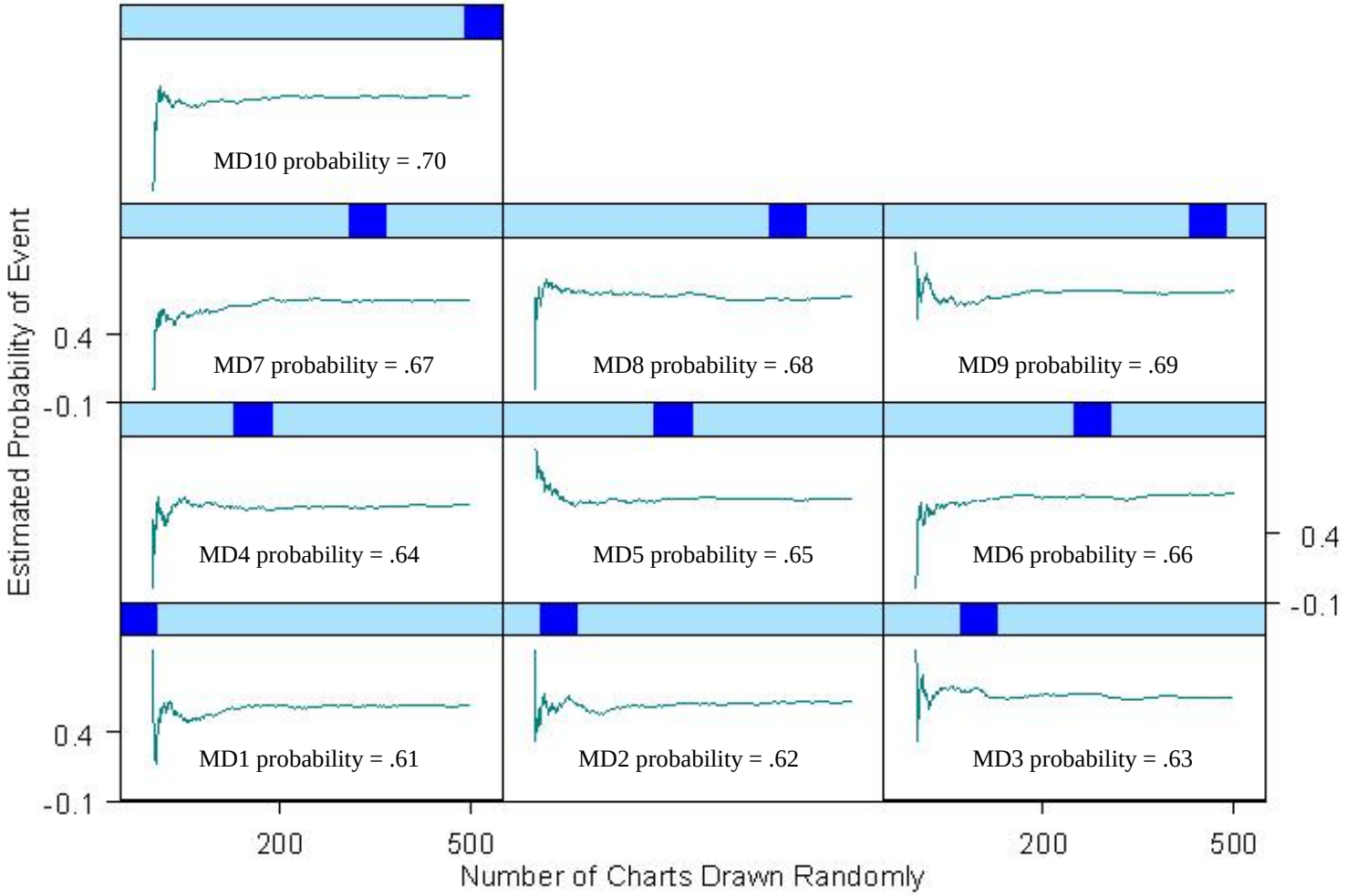
N = 500 patients


Computer Simuation of 10 Physicians each with 500 Patients
10.00
Study Probability = .894
Base Probability = .90
7.00
Study Probability = .632
Base Probability = .60
8.00
Study Probability = .698
Base Probability = .70
9.00
Study Probability = .770
Base Probability = .80
4.00
Study Probability = .406
Base Probability = .40
5.00
Study Probability = .516
Base Probability = .50
6.00
Study Probability = .580
Base Probability = .55
1.00
Study Probability = .114
Base Probability = .10
2.00
Study Probability = .198
Base Probability = .20
3.00
Study Probability = .330
Base Probability = .30
Estimated Probability of Event
0.4
-0.1
0.4
-0.1
0.4
-0.1
200
500
200
500
Number of Charts Drawn Randomly

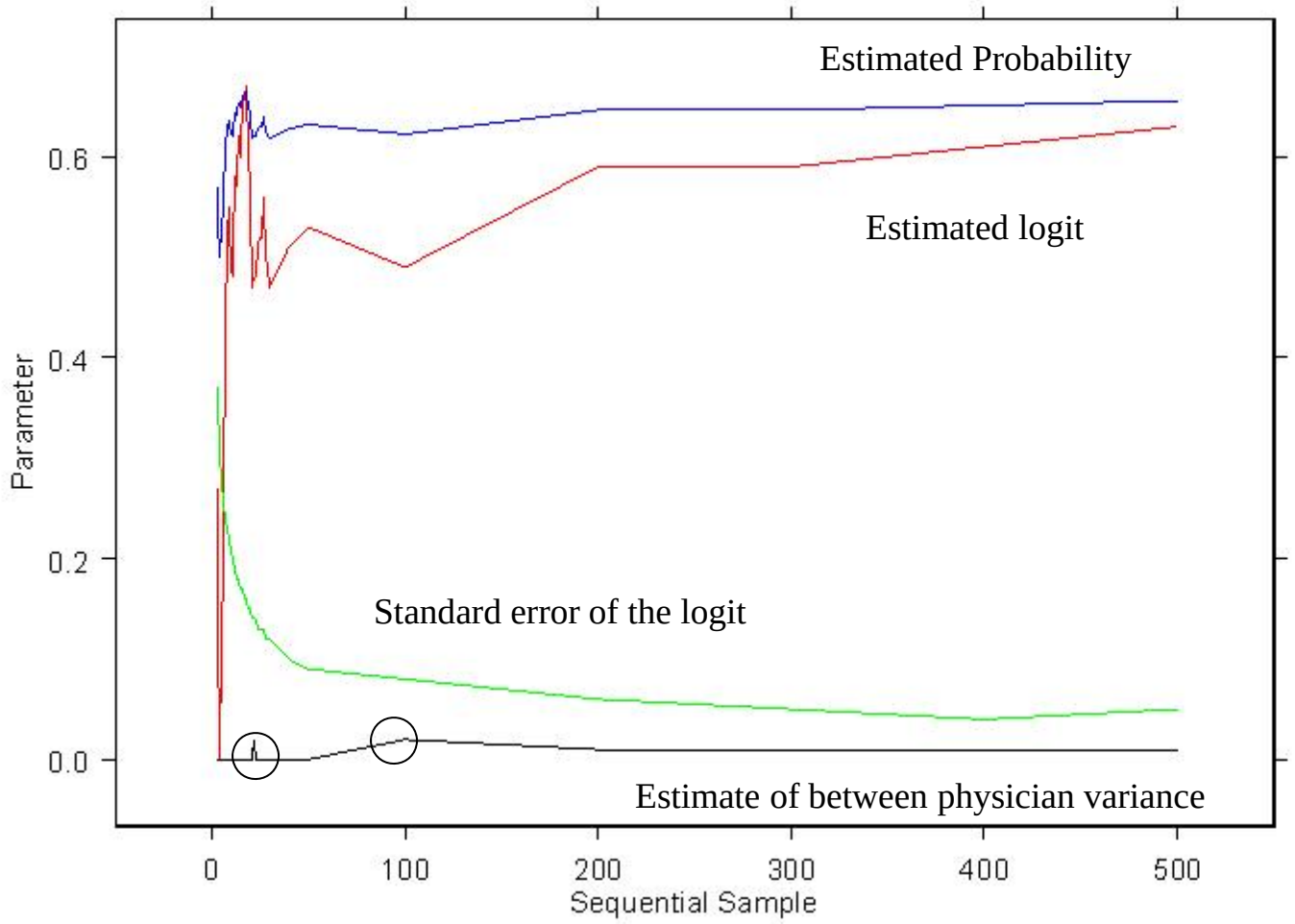

Computer Simulation Study - SET1
Estimate Parametes
0.7
0.3
-0.1
-0.5
0
5
10
15
20
25
30
Charts Sampled per Physician
Estimated Between-Physician Variance
Estimated Standard Error of the Variance
Estimated Standard Error of the Logit
Estimated Logit

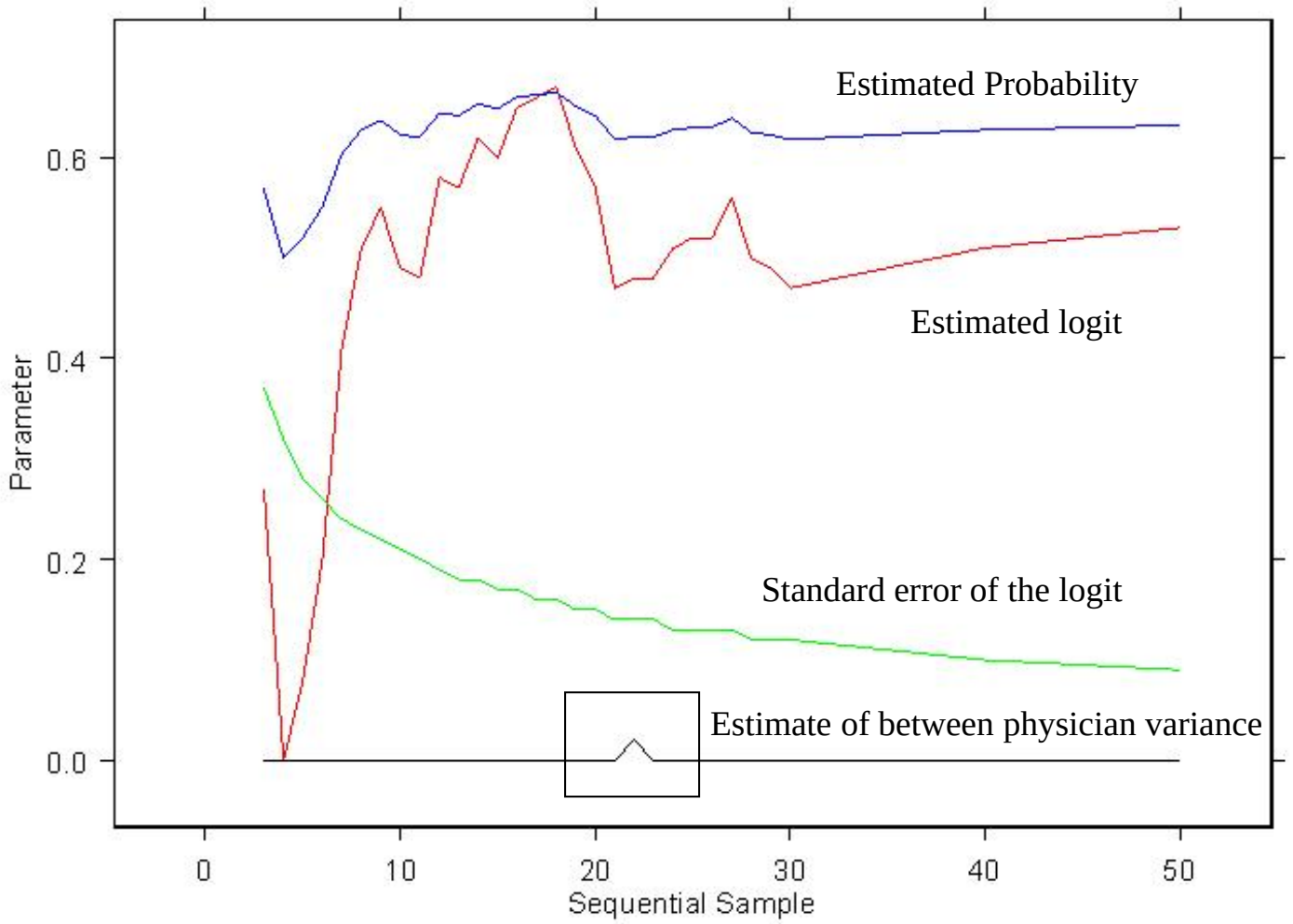

Computer Simulation Study - SET1
Probability / Reliability
0.9
0.7
0.5
0.3
0
5
10
15
20
25
30
Charts Sampled Per Physician
Estimated Reliability of the Probability
$\mathrm{Re}liability(\beta_o) = \sigma^2_{u2} \Big/ \left[\sigma^2_{u2} + \left(\sigma^2_{u1}/n_i\right)\right]$
Estimate of the Probability

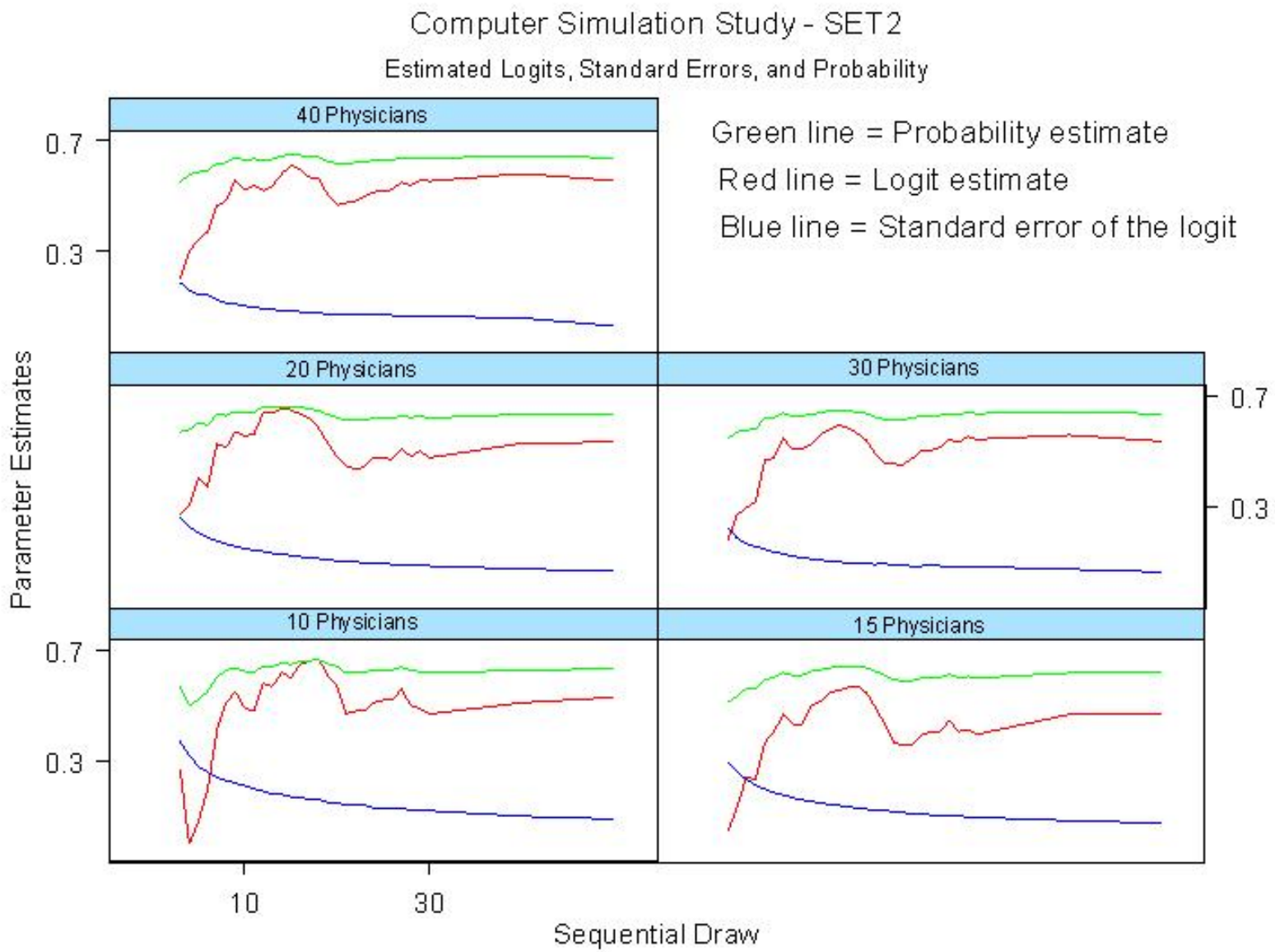
Computer Simulation Study - Set2, 10 physicians with 500 patients nested

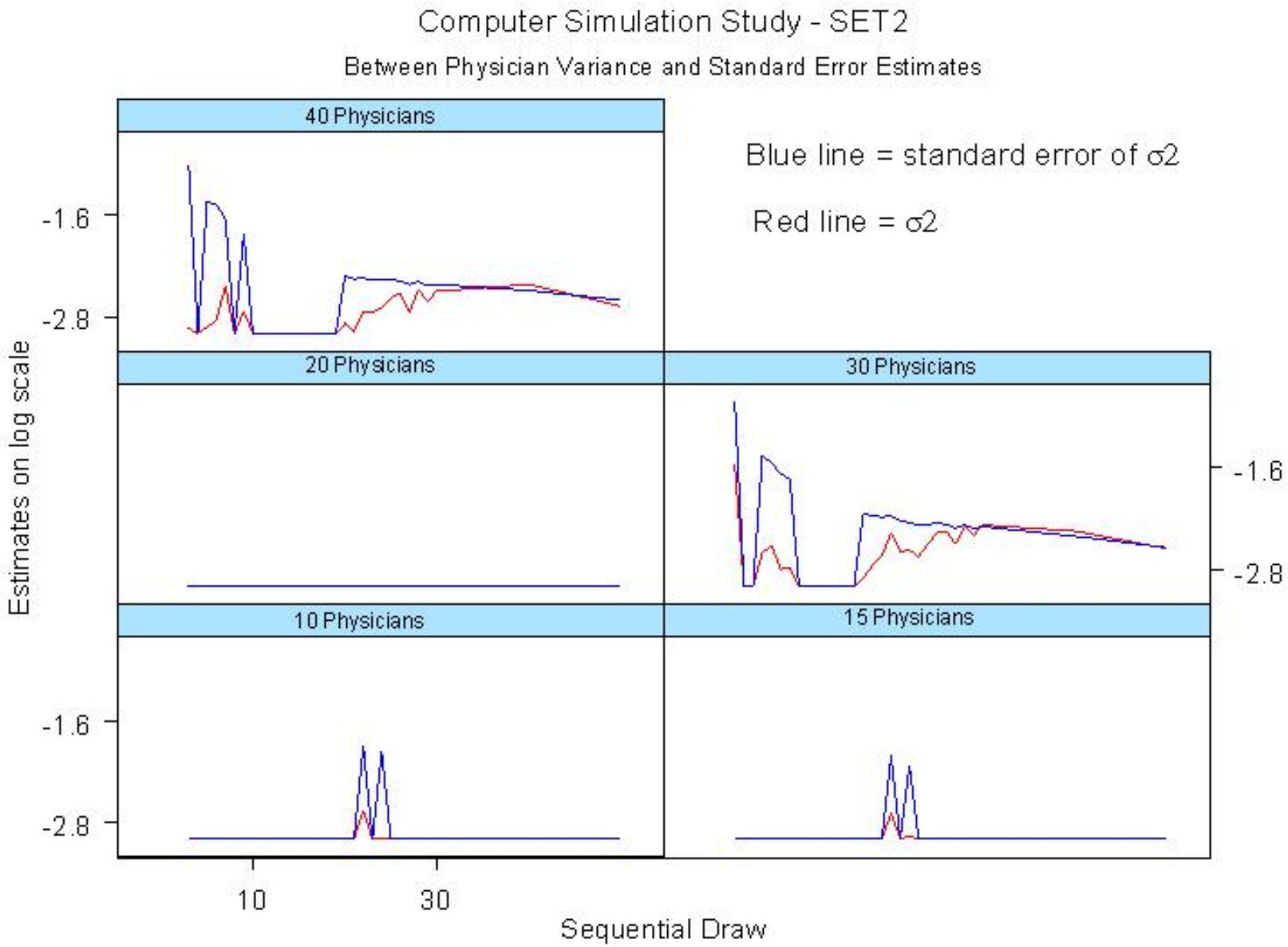


Computer Simulation - Set2




Estimated Probability
Estimated logit
Standard error of the logit
Estimate of between physician variance
Parameter
0.0
0.2
0.4
0.6
0
10
20
30
40
50
Sequential Sample


Computer Simulation Study - SET2
Estimated Logits, Standard Errors, and Probability
40 Physicians
20 Physicians
30 Physicians
10 Physicians
15 Physicians
Green line = Probability estimate
Red line = Logit estimate
Blue line = Standard error of the logit
Parameter Estimates
0.7
0.3
10
30
Sequential Draw


Computer Simulation Study - SET2
Between Physician Variance and Standard Error Estimates
40 Physicians
20 Physicians
30 Physicians
10 Physicians
15 Physicians
Blue line = standard error of σ2
Red line = σ2
Estimates on log scale
-1.6
-2.8
10
30
Sequential Draw

A look at some real data

These data are from the HEART Project
A trial to improve preventive care.

As shown in the simulation study, as physician probabilities cluster, the estimation of the between physician variability of course drops. Subsequently, operating under an assumption of highly cluster physician probabilities, estimating stable between physician variance estimates in a two-level structure will be best served by increasing the number of level two conditions (physicians).

$$\left.\begin{aligned} \text{sq7a}_{patient,\,md} &\sim \text{Binomial}(\text{denom}_{patient,\,md},\ \pi_{patient,\,md}) \\ \text{sq7a}_{patient,\,md} &= \pi_{patient,\,md} + e_{0patient,\,md}\text{bcons}^{*} \end{aligned}\right\}$$

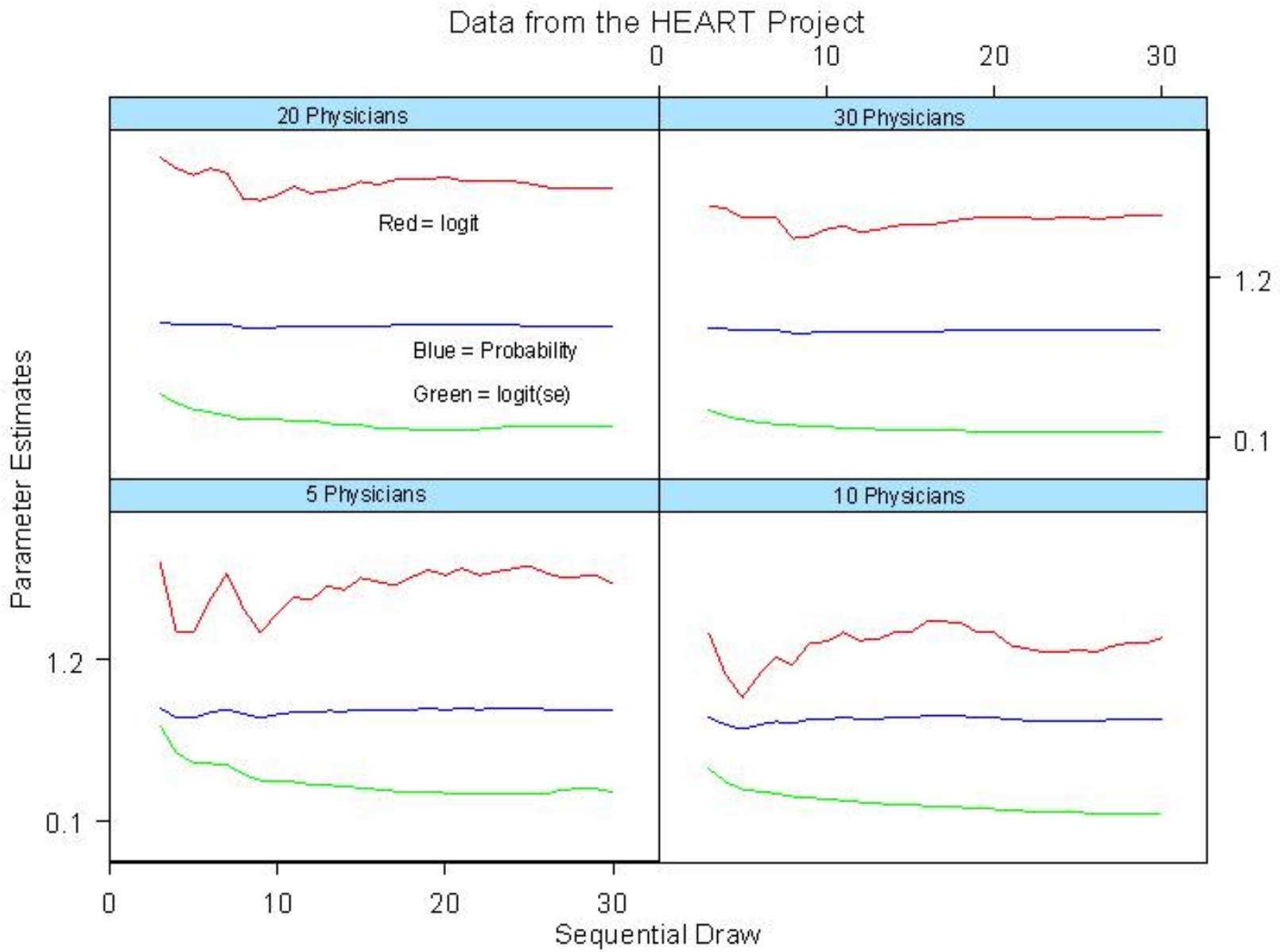
$$\text{logit}(\pi_{patient,\,md}) = \beta_{1md}\text{cons}$$

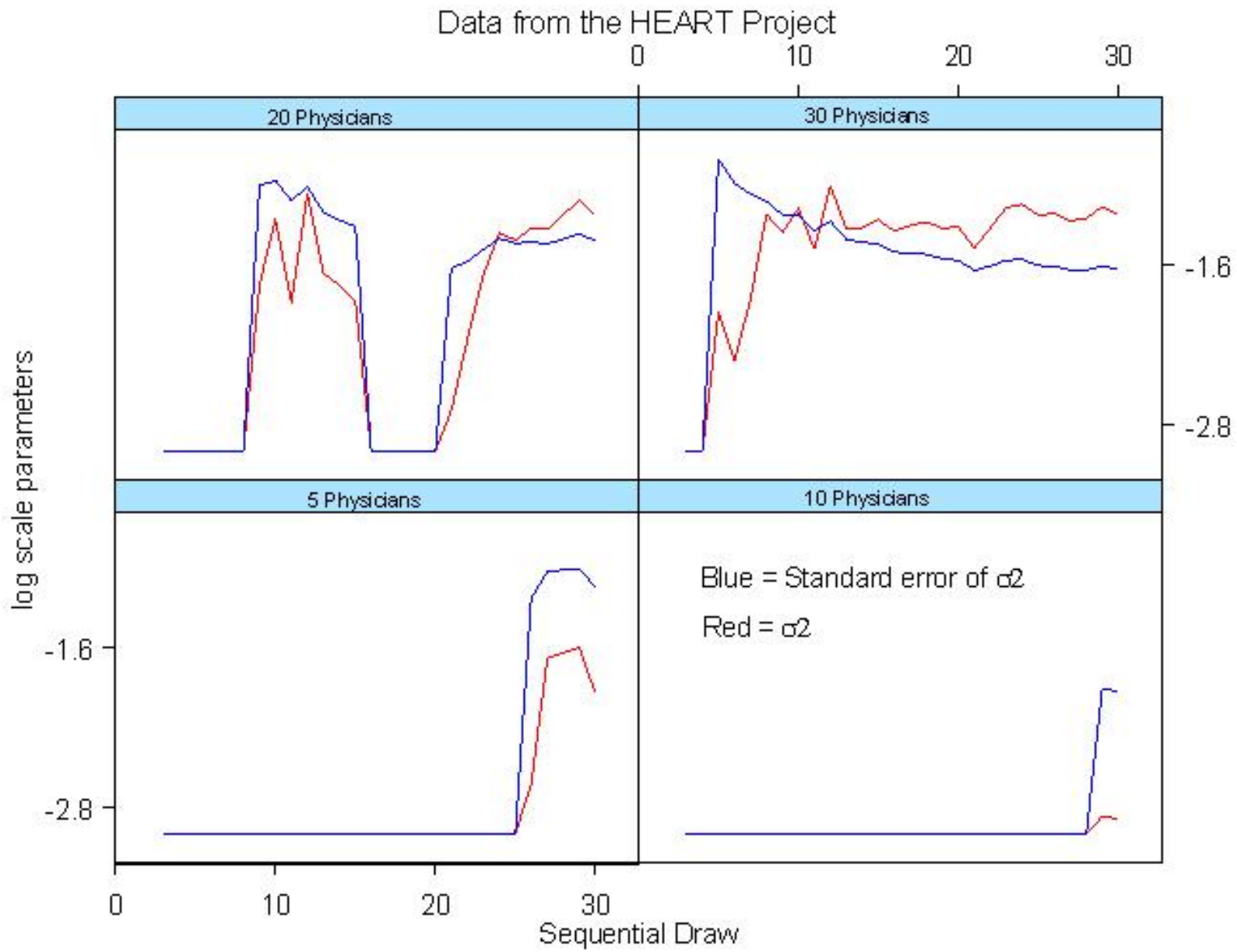
$$\beta_{1md} = 1.616(0.091) + u_{1md}$$

$$\left[u_{1md}\right] \sim \text{N}(0,\ \Omega_u) \ :\ \Omega_u = \left[0.340(0.150)\right]$$

$$\text{bcons}^{*} = \text{bcons}\left[\pi_{patient,\,md}(1 - \pi_{patient,\,md})/\text{denom}_{patient,\,md}\right]^{0.5}$$

$$\left[e_{0patient,\,md}\right] \sim (0,\ \Omega_e) \ :\ \Omega_e = \left[1.000(0.000)\right]$$


Data from the HEART Project
20 Physicians
30 Physicians
5 Physicians
10 Physicians
Red = logit
Blue = Probability
Green = logit(se)
Parameter Estimates
Sequential Draw
0
10
20
30
0.1
1.2


Data from the HEART Project
20 Physicians
30 Physicians
5 Physicians
10 Physicians
log scale parameters
-1.6
-2.8
Blue = Standard error of σ2
Red = σ2
0
10
20
30
Sequential Draw

Two-level Conditional Model

To estimate a conditional two-level model one may add either *fixed* or *random* conditions to the model providing parameter estimates and standard errors. For example if the study is concerned with a specific patient characteristic (e.g., gender of the patient) without considering it as a blocking factor or stratification issue in the study, the sampling strategy can assess stability of the influence of this factor on the model with the factor incorporated.

Here we simply expand the unconditional model by adding the patient gender parameter and allowing it to randomly vary across attending physicians. Subsequently, patient gender is a *random* variable in the model with an estimated variance of gender across the higher levels (physician).

$$\text{logit}\left(\pi_{ij}\right) = \beta_{oj} X_{o} + \beta_{1j} gender_{1}$$

$$\beta_{oj} = \beta_{o} + u_{oj}$$

$$\beta_{1j} = \beta_{1} + u_{1j}$$

where $u_{oj} \approx N\left(0, \sigma^2_{uo2}\right)$ $u_{1j} \approx N\left(0, \sigma^2_{u12}\right)$

$$\left.\begin{aligned} y_{ij} &\sim \text{Binomial}(n_{ij}, \pi_{ij}) \\ y_{ij} &= \pi_{ij} + e_{0ij}x^{*}_{0} \end{aligned}\right\}$$

$$\text{logit}(\pi_{ij}) = \beta_{1j}x_{1} + \beta_{2j}x_{2ij}$$

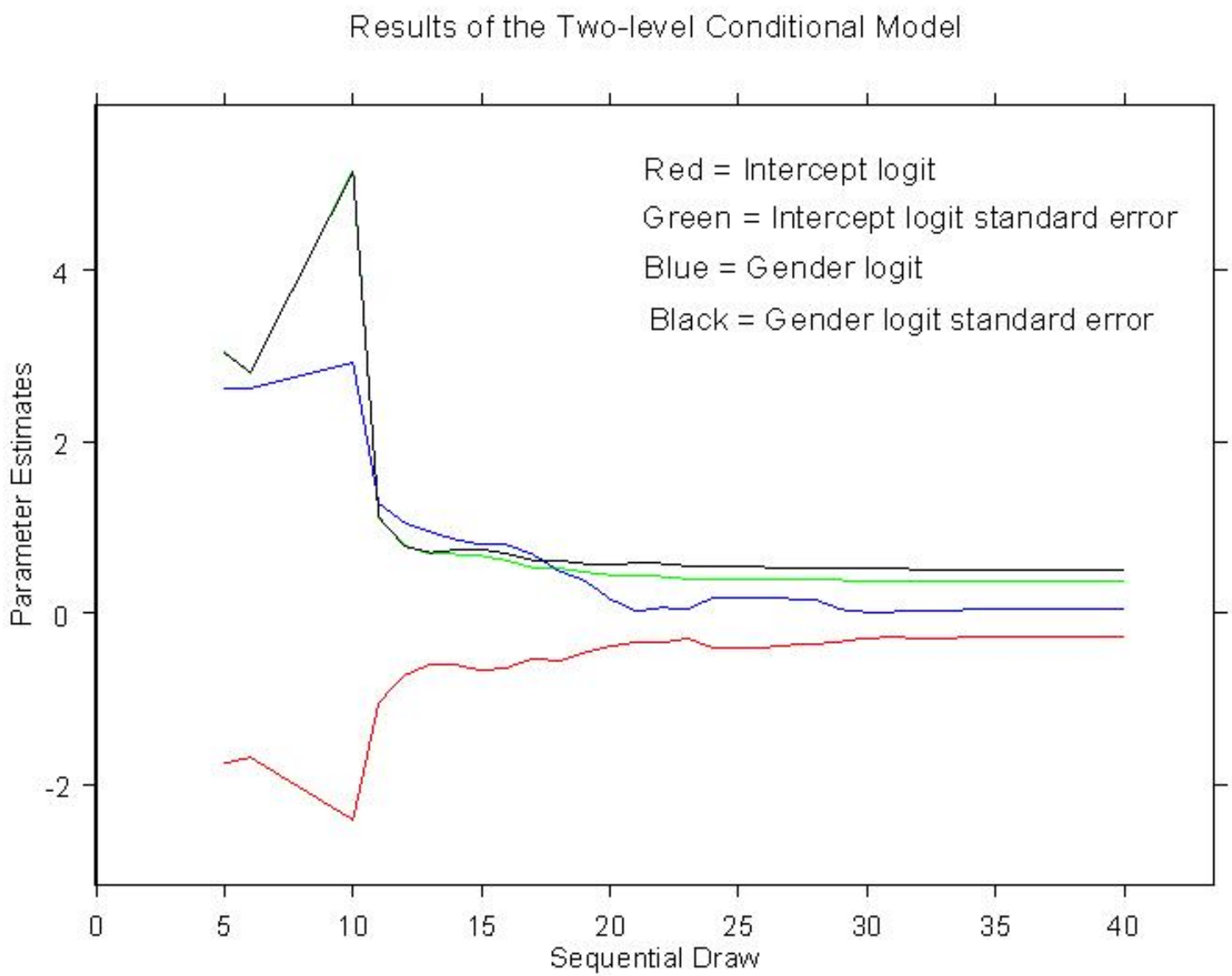
$$\beta_{1j} = \beta_{1} + u_{1j}$$

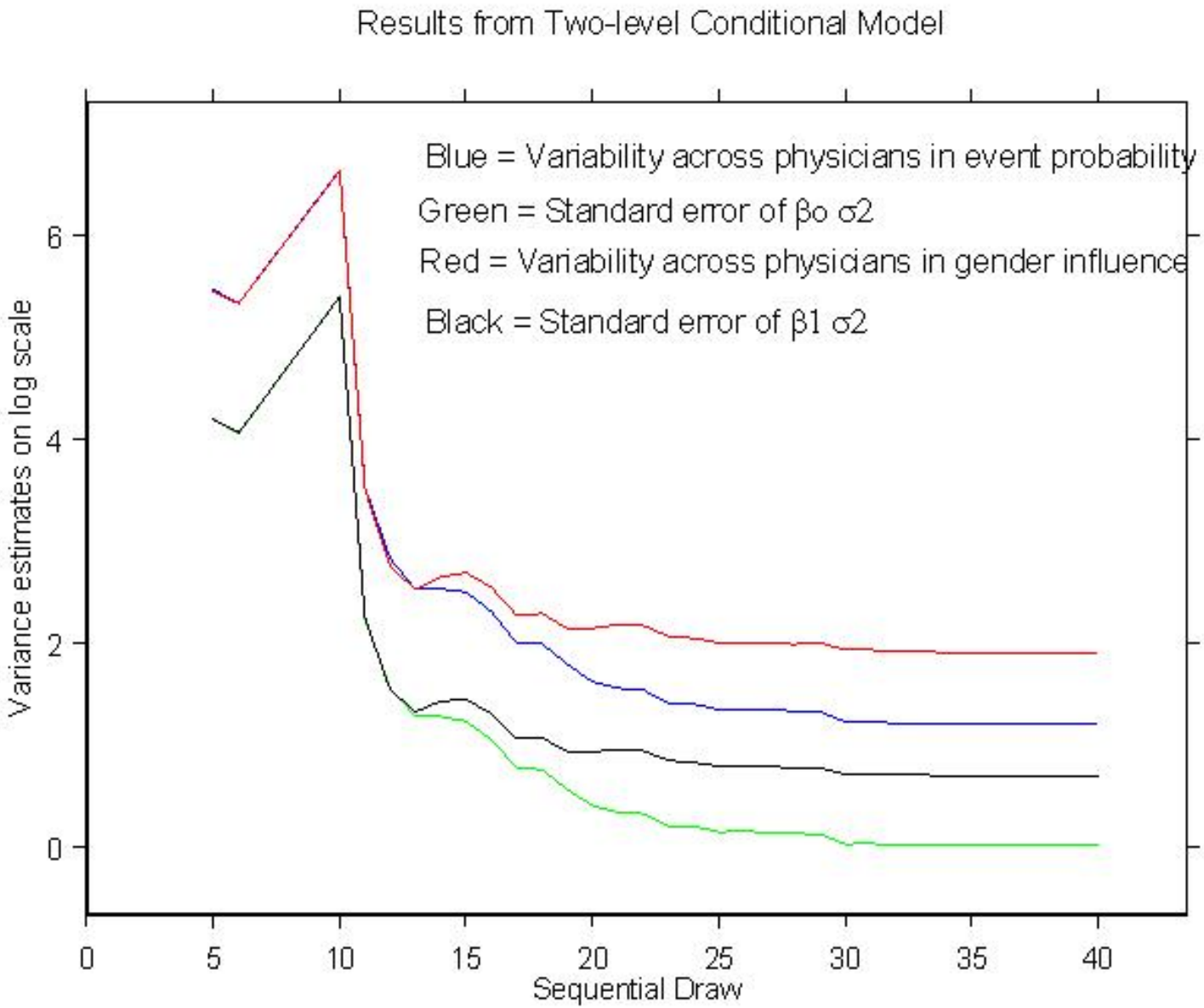
$$\beta_{2j} = \beta_{2} + u_{2j}$$

$$\begin{bmatrix} u_{1j} \\ u_{2j} \end{bmatrix} \sim \text{N}(0,\ \Omega_{u}) \ :\ \Omega_{u} = \begin{bmatrix} \sigma^{2}_{u1} & \\ \sigma_{u21} & \sigma^{2}_{u2} \end{bmatrix}$$

$$x^{*}_{0} = x_{0}[\pi_{ij}(1 - \pi_{ij})/n_{ij}]^{0.5}$$

$$\begin{bmatrix} e_{0ij} \end{bmatrix} \sim (0,\ \Omega_{e}) \ :\ \Omega_{e} = \begin{bmatrix} 1 \end{bmatrix}$$


Results of the Two-level Conditional Model
Red = Intercept logit
Green = Intercept logit standard error
Blue = Gender logit
Black = Gender logit standard error
Parameter Estimates
4
2
0
-2
0
5
10
15
20
25
30
35
40
Sequential Draw


Results from Two-level Conditional Model
Blue = Variability across physicians in event probability
Green = Standard error of βo σ2
Red = Variability across physicians in gender influence
Black = Standard error of β1 σ2
Variance estimates on log scale
0
2
4
6
0
5
10
15
20
25
30
35
40
Sequential Draw

$$\left.\begin{aligned} sq5_{patient,\,md,\,clinic} &\sim \text{Binomial}(denom_{patient,\,md,\,clinic},\ \pi_{patient,\,md,\,clinic}) \\ sq5_{patient,\,md,\,clinic} &= \pi_{patient,\,md,\,clinic} + e_{0patient,\,md,\,clinic}\text{bcons}^{*} \end{aligned}\right\}$$

$$\text{logit}(\pi_{patient,\,md,\,clinic}) = \beta_{1md,\,clinic}\text{cons}$$

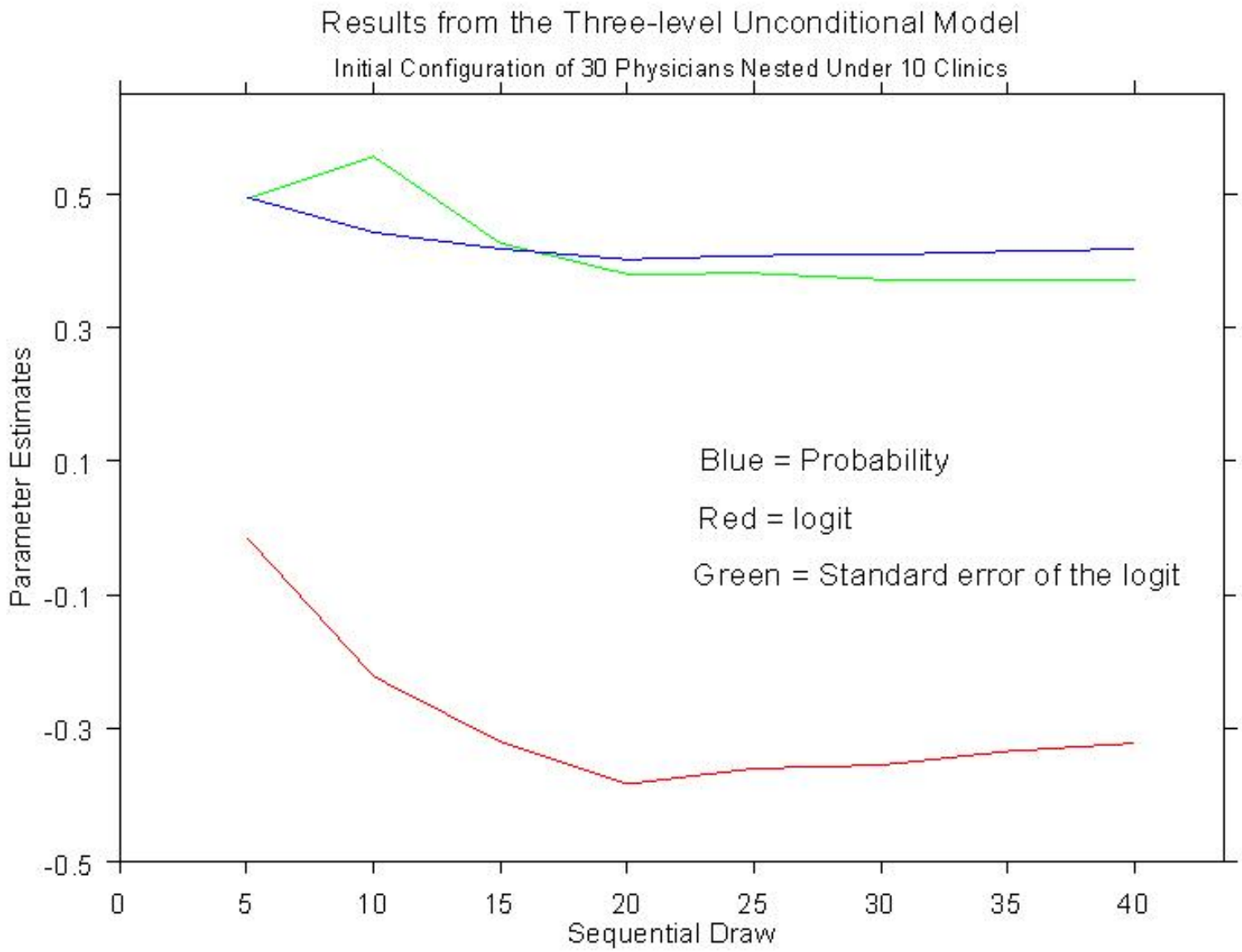
$$\beta_{1md,\,clinic} = \beta_1 + v_{1clinic} + u_{1md,\,clinic}$$

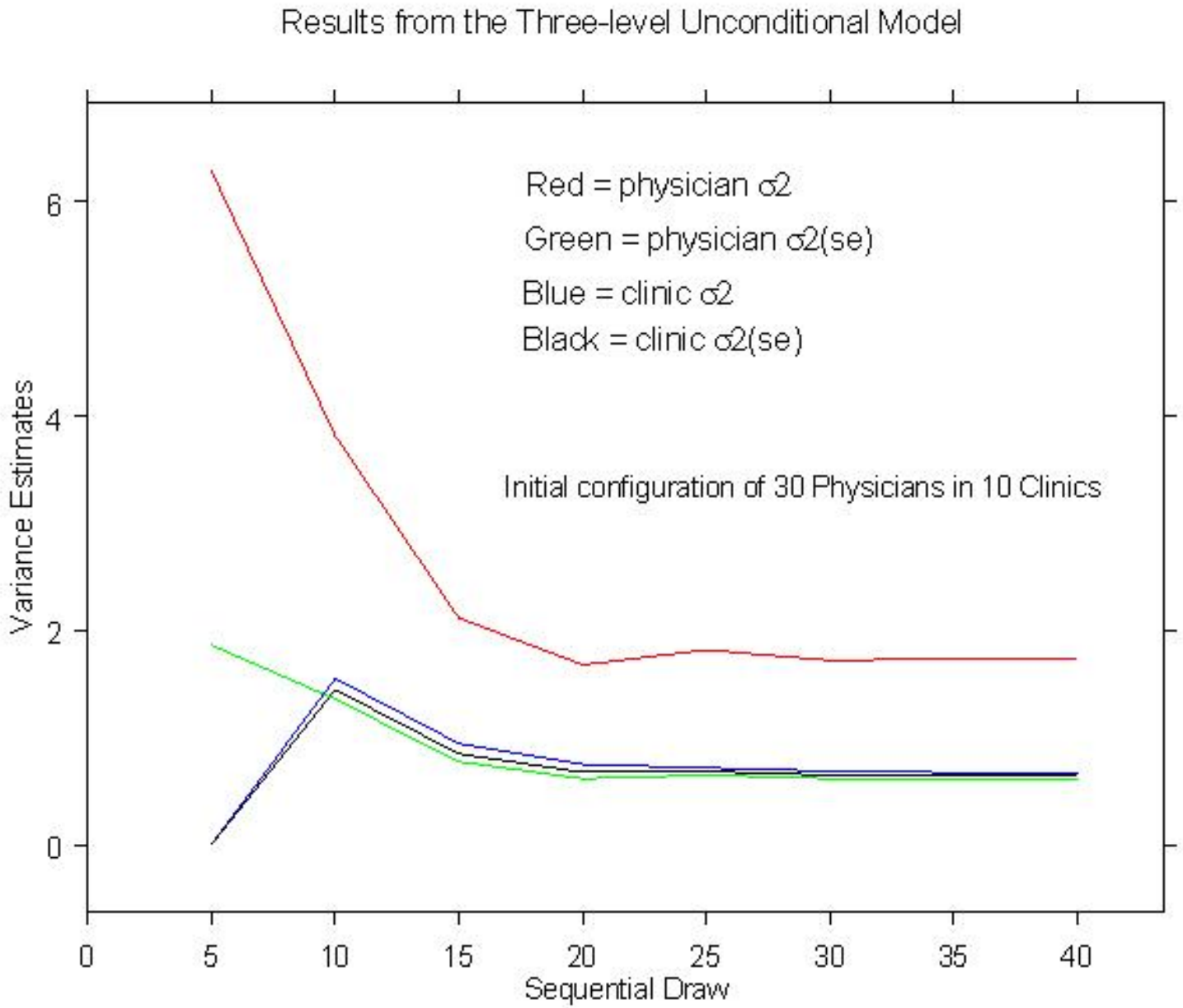
$$\begin{bmatrix} v_{1clinic} \end{bmatrix} \sim N(0,\ \Omega_v) \ :\ \Omega_v = \begin{bmatrix} \sigma^2_{v\,1} \end{bmatrix}$$

$$\begin{bmatrix} u_{1md,\,clinic} \end{bmatrix} \sim N(0,\ \Omega_u) \ :\ \Omega_u = \begin{bmatrix} \sigma^2_{u\,1} \end{bmatrix}$$

$$\text{bcons}^{*} = \text{bcons}\left[\pi_{patient,\,md,\,clinic}(1 - \pi_{patient,\,md,\,clinic})/denom_{patient,\,md,\,clinic}\right]^{0.5}$$

$$\begin{bmatrix} e_{0patient,\,md,\,clinic} \end{bmatrix} \sim (0,\ \Omega_e) \ :\ \Omega_e = \begin{bmatrix} \sigma^2_{e\,0} \end{bmatrix}$$


Results from the Three-level Unconditional Model
Initial Configuration of 30 Physicians Nested Under 10 Clinics
Parameter Estimates
0.5
0.3
0.1
-0.1
-0.3
-0.5
0
5
10
15
20
25
30
35
40
Sequential Draw
Blue = Probability
Red = logit
Green = Standard error of the logit


Results from the Three-level Unconditional Model
Red = physician σ2
Green = physician σ2(se)
Blue = clinic σ2
Black = clinic σ2(se)
Initial configuration of 30 Physicians in 10 Clinics
Variance Estimates
Sequential Draw
0
2
4
6
0
5
10
15
20
25
30
35
40

$$\left.\begin{aligned} sq5_{patient,md,clinic} &\sim \text{Binomial}(denom_{patient,md,clinic}, \pi_{patient,md,clinic}) \\ sq5_{patient,md,clinic} &= \pi_{patient,md,clinic} + e_{0patient,md,clinic} bcons^{*} \end{aligned}\right\}$$

$$\text{logit}(\pi_{patient,md,clinic}) = \beta_{1md,clinic}\text{cons} + \beta_{2md,clinic}\text{gender}_{patient,md,clinic}$$

$$\beta_{1md,clinic} = \beta_1 + v_{1clinic} + u_{1md,clinic}$$

$$\beta_{2md,clinic} = \beta_2 + v_{2clinic} + u_{2md,clinic}$$

$$\begin{bmatrix} v_{1clinic} \\ v_{2clinic} \end{bmatrix} \sim N(0,\ \Omega_v) \ :\ \Omega_v = \begin{bmatrix} \sigma^2_{v1} & \\ \sigma_{v21} & \sigma^2_{v2} \end{bmatrix}$$

$$\begin{bmatrix} u_{1md,clinic} \\ u_{2md,clinic} \end{bmatrix} \sim N(0,\ \Omega_u) \ :\ \Omega_u = \begin{bmatrix} \sigma^2_{u1} & \\ \sigma_{u21} & \sigma^2_{u2} \end{bmatrix}$$

$$bcons^{*} = bcons[\pi_{patient,md,clinic}(1 - \pi_{patient,md,clinic})/denom_{patient,md,clinic}]^{0.5}$$

$$\begin{bmatrix} e_{0patient,md,clinic} \end{bmatrix} \sim (0,\ \Omega_e) \ :\ \Omega_e = \begin{bmatrix} 1 \end{bmatrix}$$

To estimate power of intervention effects in hierarchical models the researcher needs to have some knowledge beforehand about the relevant population parameters. As Goldstein has indicated, pilot studies or even simulation studies are required to optimize the design of a multilevel study (1987, p. 87).

To date, little work has been done on establishing power/sample size estimation equations for intervention trials in multilevel studies. Some work by Snijders and Bosker on two-level standard error stability has been done, but not dealing with intervention effects. Our studies will provide the initial framework for establishing this research in hierarchical primary care intervention trials.