Improving the SRTR Methodology Used to Identify Potentially Underperforming Transplant Programs in the United States

Nicholas Salkowski, PhD*[1]
Erick Edwards, PhD[2]
Tabitha Leighton, MPH[1]
Ajay Israni, MD MS[1]
Bertram Kasiske, MD[1]
and Jon Snyder, PhD[1]

May 22, 2013

[1]Scientific Registry of Transplant Recipients, Minneapolis Medical Research Foundation, Minneapolis, MN, United States
and
[2]United Network for Organ Sharing, Richmond, VA, United States.

:SRTR
SCIENTIFIC REGISTRY OF TRANSPLANT RECIPIENTS

*Presenter

Speaker Disclosure

Nicholas Salkowski, PhD

Biostatistician
Scientific Registry of Transplant Recipients, Minneapolis Medical Research Foundation, Minneapolis, MN, United States

- I have no financial relationships to disclose within the past 12 months relevant to my presentation.

- My presentation does not include discussion of off-label or investigational use.

- I do not intend to reference unlabeled/unapproved uses of drugs or products in my presentation.

Why explore a Bayesian Framework?

Hierarchical Bayesian approaches are favored over the current statistical approach in assessing provider performance.

- **COPSS Report**

STATISTICAL ISSUES IN ASSESSING HOSPITAL PERFORMANCE

Commissioned by the Committee of Presidents of Statistical Societies

The COPSS-CMS White Paper Committee:

Arlene S. Ash, PhD; Stephen E. Fienberg, PhD; Thomas A. Louis, PhD
Sharon-Lise T. Normand, PhD; Thérèse A. Stukel, PhD; Jessica Utts, PhD

Original report submitted to CMS on November 28, 2011
Revised on January 27, 2012

- **2012 OPTN/SRTR Consensus Conference on Transplant Program Quality and Surveillance**

- **SRTR Technical Advisory Committee (STAC)**

What questions are answered by the different approaches?

Current:

Is a particular center performing **as expected**?

- Produces a Yes/No Decision

- P Value shows how extreme the data would be if the center had average performance

Bayesian:

What is the **probability** that a particular center is **underperforming**?

- Produces a probability

- Produces a distribution for center performance

PSR Output Comparison: Small Center "A"

- Number of Transplants: 6
- Observed 1-Year Patient Deaths: **1**
- Expected 1-Year Patient Deaths: **0.18**
- **O/E Ratio: 5.42**
- O - E: **0.82**
- 95% Confidence Interval: **(0.14, 30.20)**
- Two-sided p-value: **0.337**
- One-sided p-value: **0.168**
- Current Flag: **TRUE**

95% Credible Interval
0.28   3.31

1.37

0   1   2   3   4

Hazard Ratio (HR)
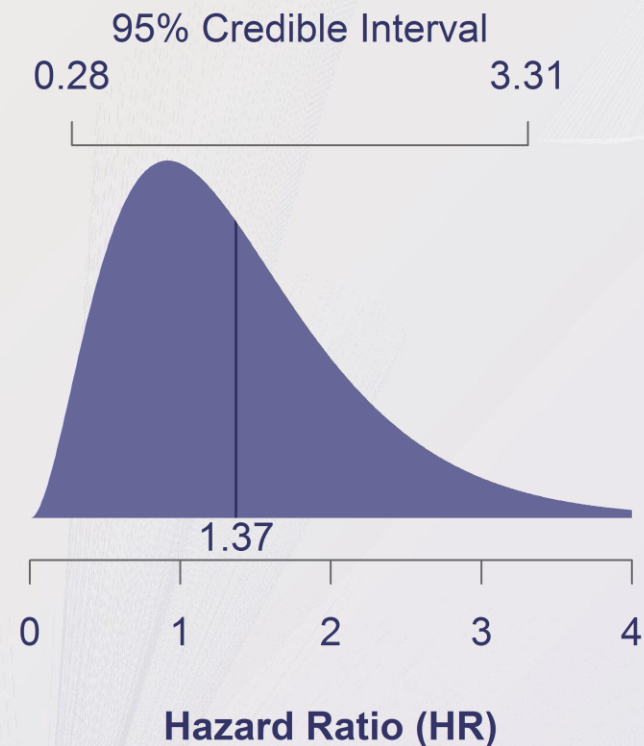
New Simulations

The SRTR simulated data:

- Simulated patient survival for all Heart, Kidney, Liver, and Lung programs with expected 1-year adult patient deaths in the July 2012 PSR cohort, assuming that each program was performing as expected 2,500 times.

- Simulated patient survival for the same programs, assuming that their patient death rates were 2 times their expected rates 2,500 times.

The SRTR then calculated the flagging rates for 57,915 possible Bayesian flagging algorithms, then calculated the **score** for **each algorithm!**

Evaluation Score Formula

Afterwards, the SRTR developed the following flagging algorithm score:

- Penalty of 5 points for every percent the false positive rate differed from 5% for each program when the simulated HR = 1.0

- Penalty of 1 point for every percent the true positive rate differed from 100% for each program when the simulated HR = 2.0
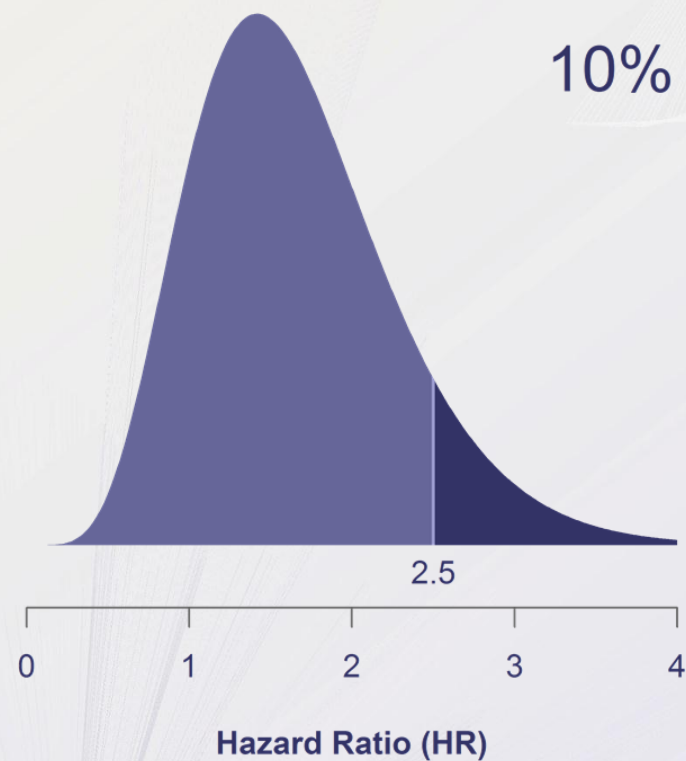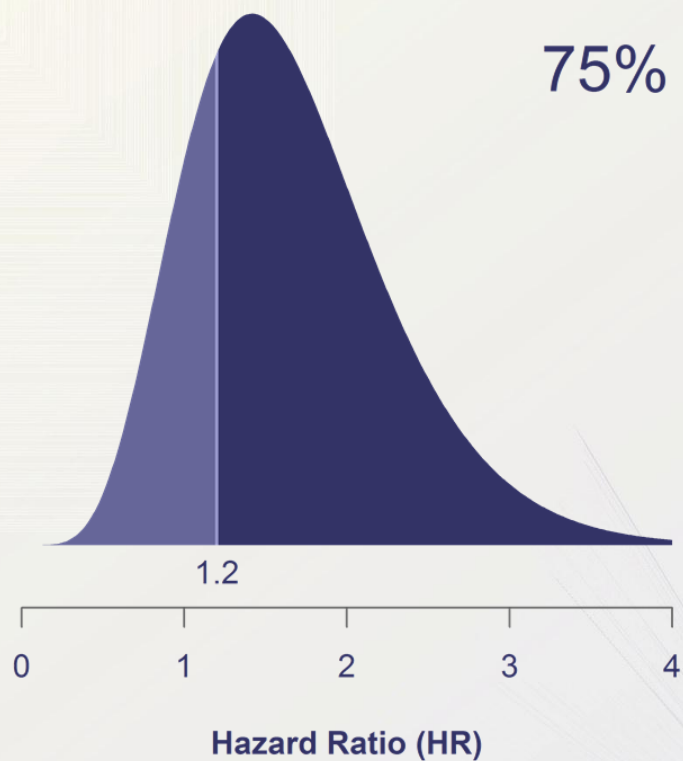
- Low scores are better

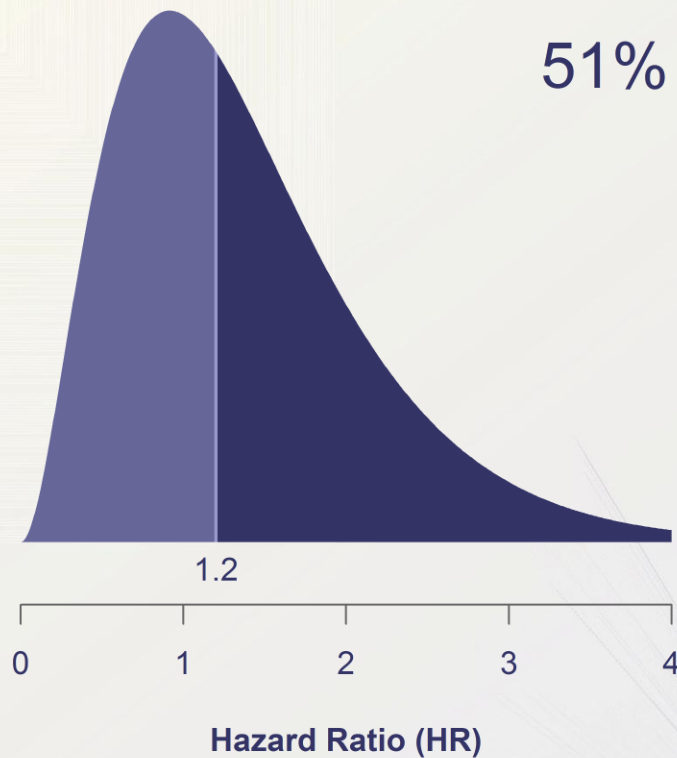# Results

Distribution of Algorithm Scores (93.7% < 800)
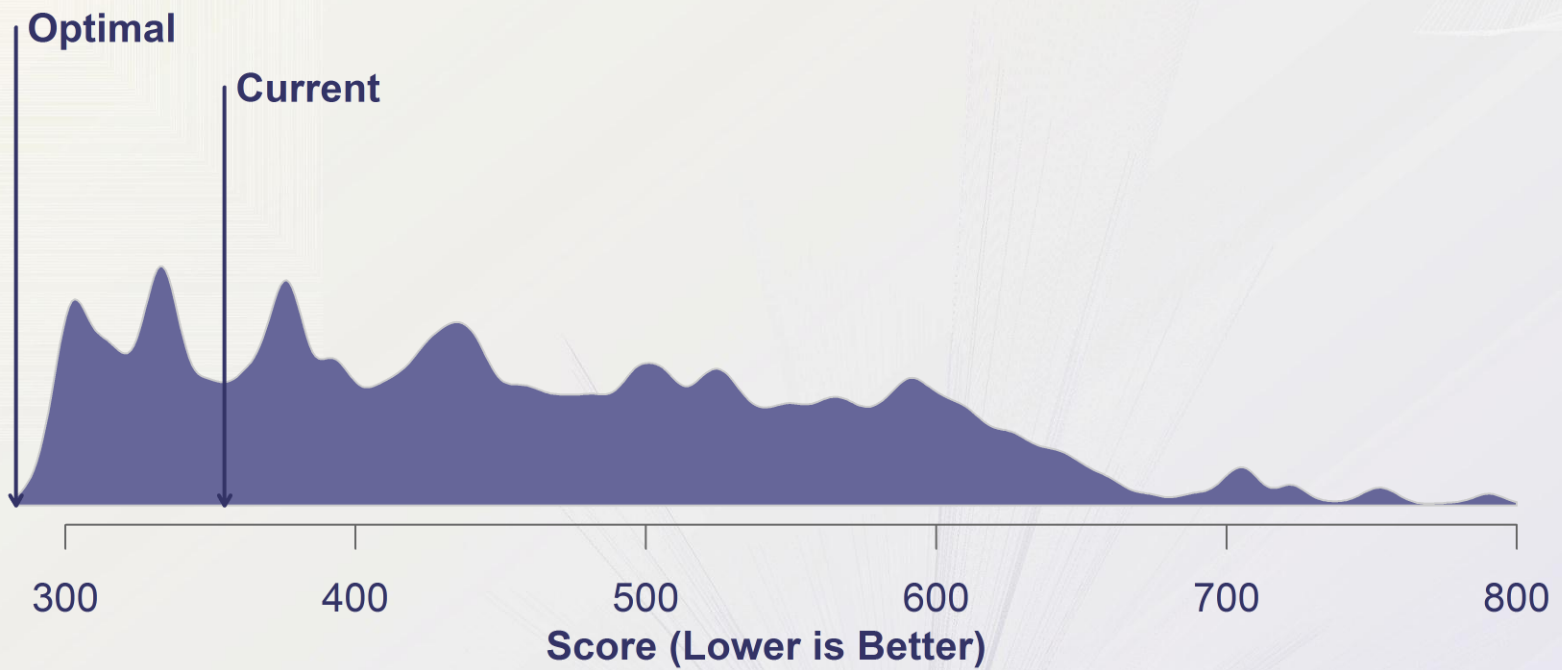
Proposed MPSC Flagging Criteria
April 10, 2013

Bayesian PSR Output Example:
Small Center "A" -- No Flag
[But Close for Criterion #2!]

51%

9%

1.2

2.5

0	1	2	3	4

Hazard Ratio (HR)

Hazard Ratio (HR)

January 2008 Flagged Programs By Volume

Volume	Positive	Transplants	Current	Proposed MPSC Bayesian Criteria
[1,10)	FALSE	268	60	18
[10,50)	FALSE	159	6	6
[50,100)	FALSE	186	3	3
[100,250)	FALSE	374	2	2

Volume	Positive	Transplants	Current	Proposed MPSC Bayesian Criteria
[1,10)	TRUE	96	17	6
[10,50)	TRUE	241	10	10
[50,100)	TRUE	1,032	14	14
[100,250)	TRUE	1,770	11	11
[250,600]	TRUE	1,850	5	5

Only the numbers for small volume programs differ.

NOTE: This table ONLY contains data for programs flagged using the Current Flag algorithm. Programs that would have been flagged ONLY by alternative algorithms are not included.

July 2012 PSR Cohort By Volume

Volume	Programs	Transplants	Current	Proposed MPSC Bayesian Criteria
[1,10)	223	799	54	15
[10,50)	270	7,519	22	44
[50,100)	126	9,139	11	19
[100,250)	147	23,694	11	15
[250,744]	61	23,977	4	4
Total	827	65,128	102	97

Other simulations show Bayesian flagging algorithms should . . .

- Have fewer false positive flags for small volume programs
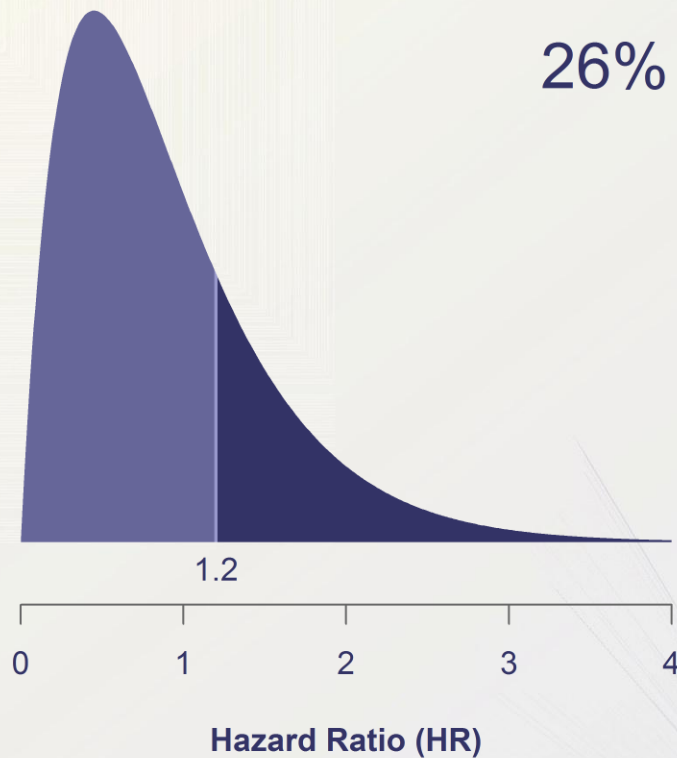- Have more true positive flags for larger volume programs

12

Supplemental Slides

PSR Output Comparison: Small Center "B"

- Number of Transplants: 6
- Observed 1-Year Patient Deaths: **0**
- Expected 1-Year Patient Deaths: **0.22**
- **O/E Ratio: 0**
- O - E: **-0.22**
- 95% Confidence Interval: **(0.00, 13.75)**
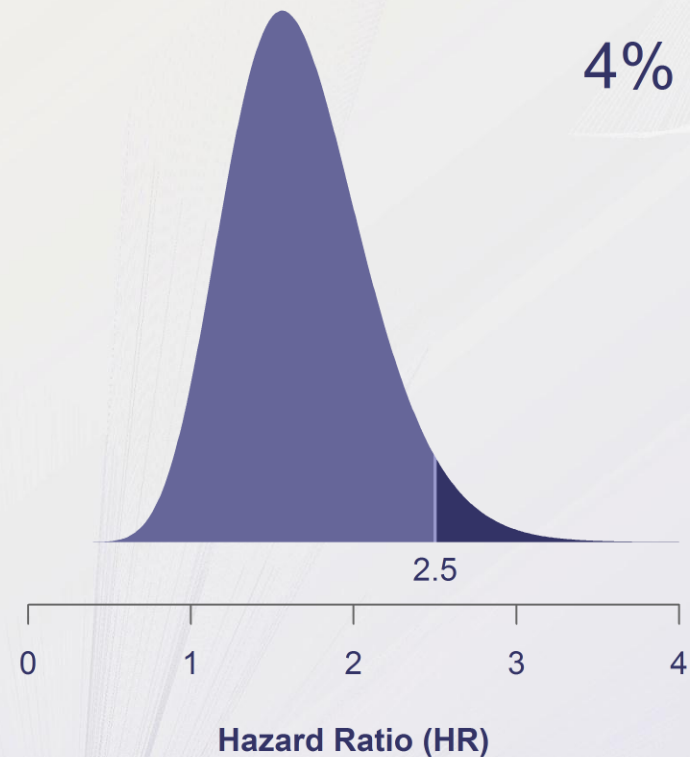- Two-sided p-value: **0.999**
- Current Flag: **FALSE**

95% Credible Interval
0.11   2.51

0.9

0   1   2   3   4

Hazard Ratio (HR)
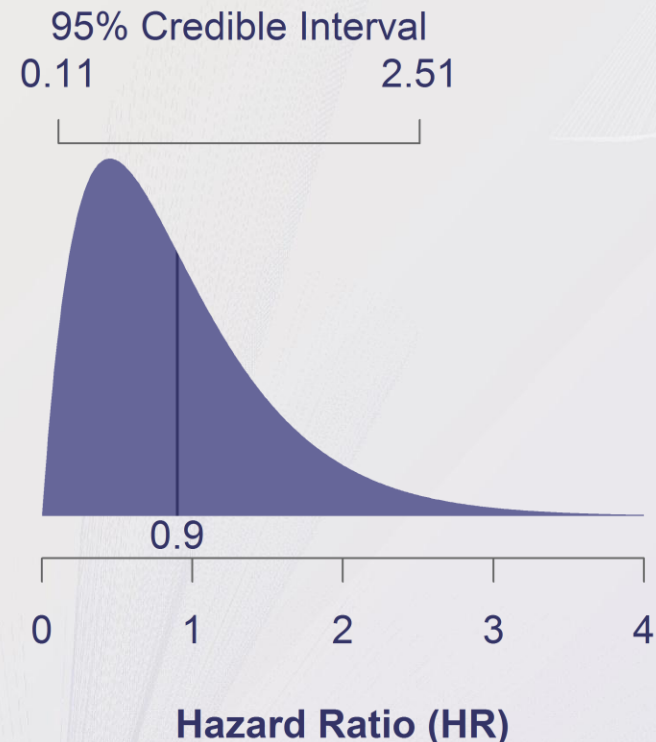
Bayesian PSR Output Example: Small Center "B" -- No Flag

26%

3%

1.2

0   1   2   3   4

Hazard Ratio (HR)
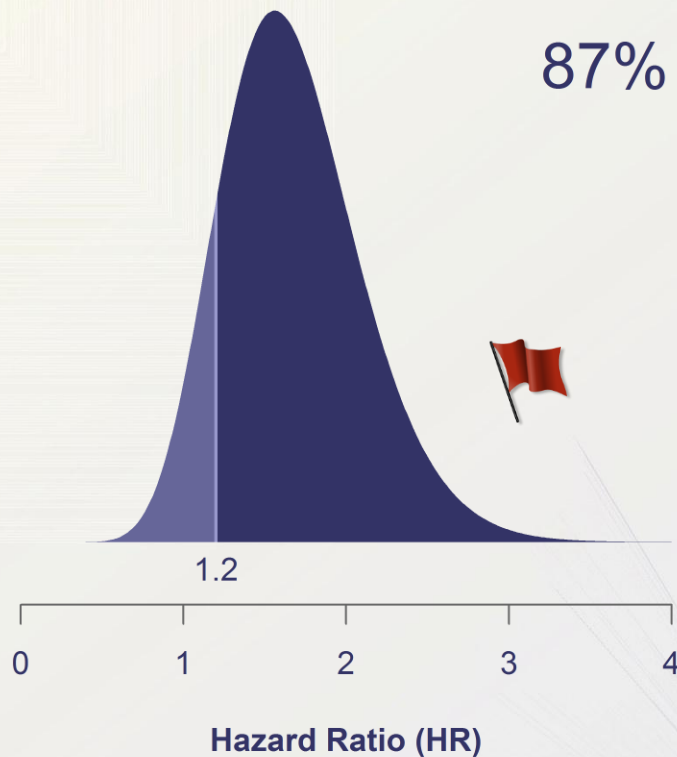
2.5

0   1   2   3   4

Hazard Ratio (HR)
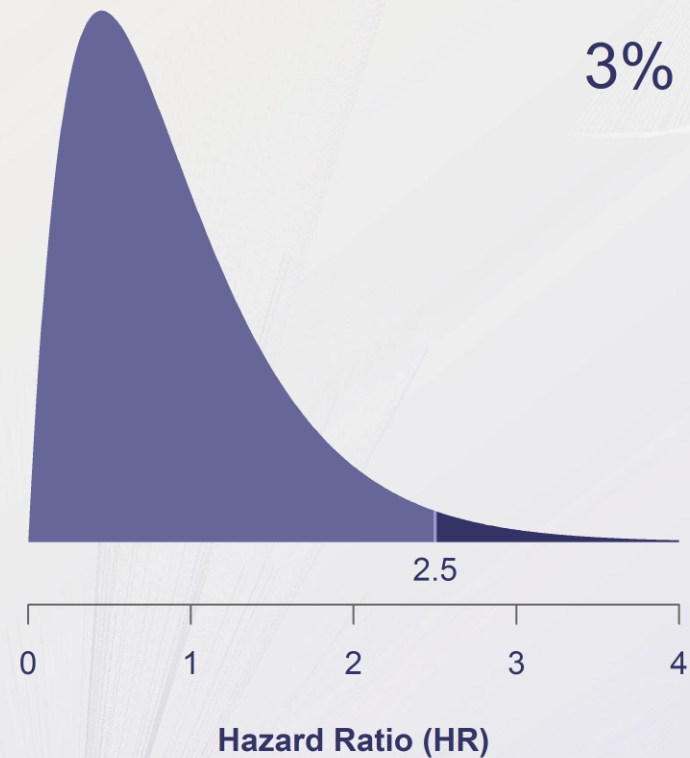
PSR Output Comparison: Large Center "C"

- Number of Transplants: **299**
- Observed 1-Year Patient Deaths: **13**
- Expected 1-Year Patient Deaths: **6.97**
- O/E Ratio: **1.87**
- O - E: **6.03**
- 95% Confidence Interval: **(0.99, 3.19)**
- Two-sided p-value: **0.052**
- One-sided p-value: **0.026**
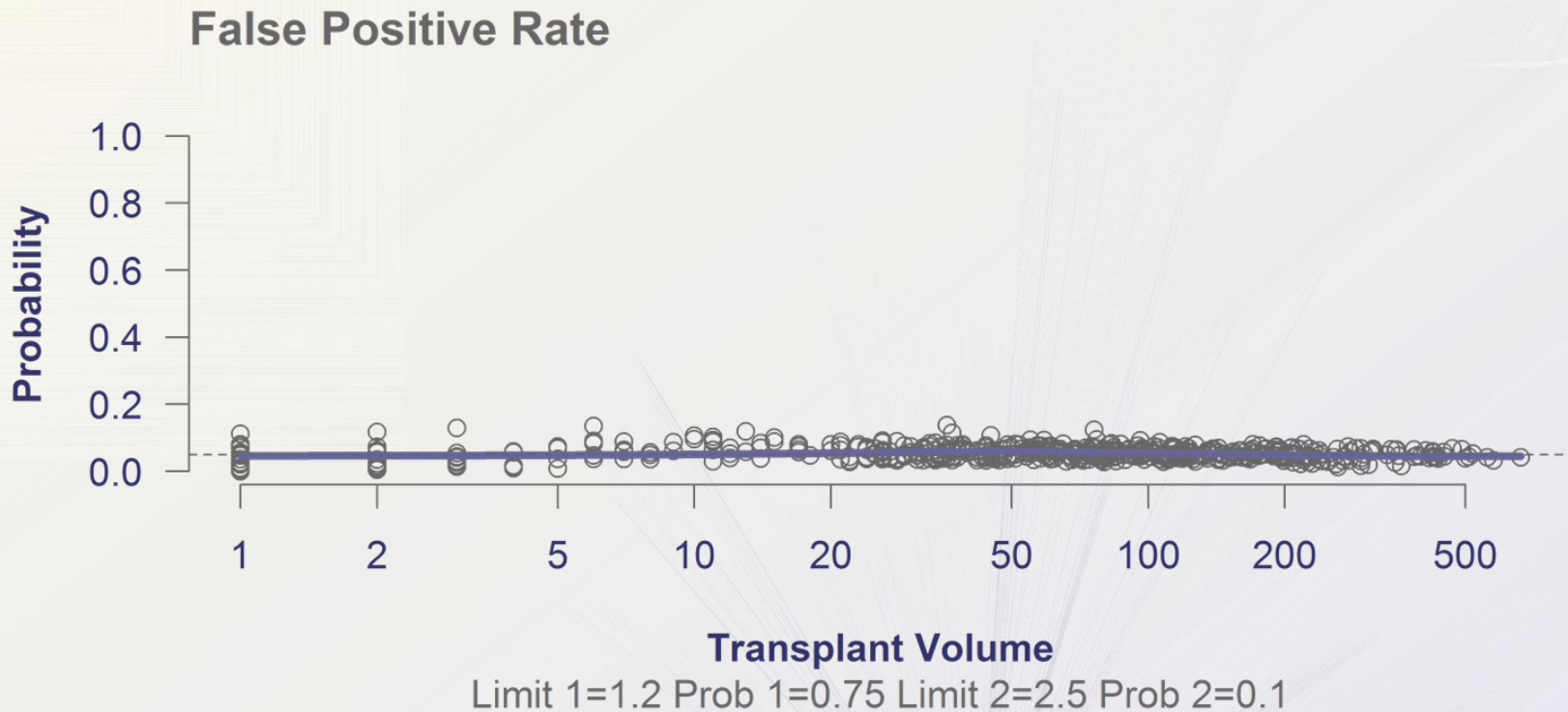- Current Flag: **TRUE**

95% Credible Interval
0.94   2.62

1.67

0   1   2   3   4

Hazard Ratio (HR)
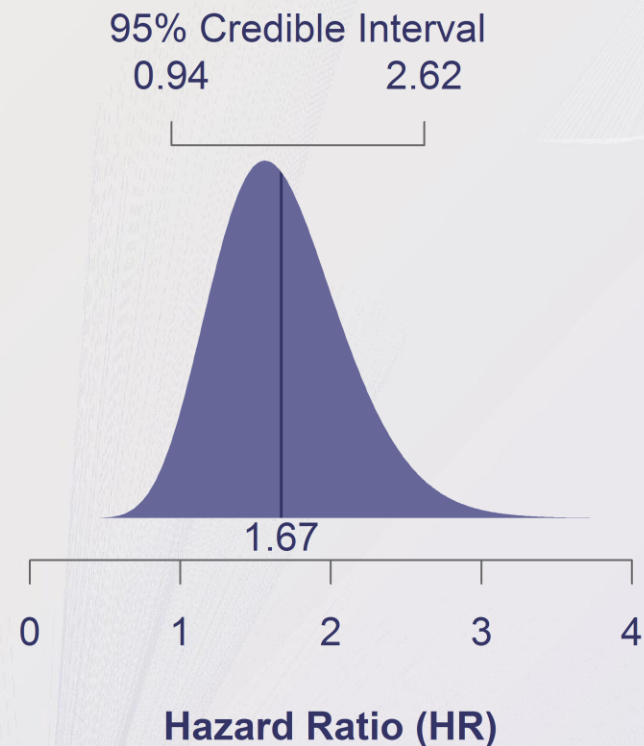
Bayesian PSR Output Example:
Large Center "C" -- Flagged Under Criterion #1

87%

4%

1.2

2.5

0 1 2 3 4

0 1 2 3 4

Hazard Ratio (HR)

Hazard Ratio (HR)

Proposed MPSC Flag
False Positive Rate: All

False Positive Rate

Limit 1=1.2 Prob 1=0.75 Limit 2=2.5 Prob 2=0.1

Proposed MPSC Flag
True Positive Rate: All

True Positive Rate

Probability vs Transplant Volume

Limit 1=1.2 Prob 1=0.75 Limit 2=2.5 Prob 2=0.1
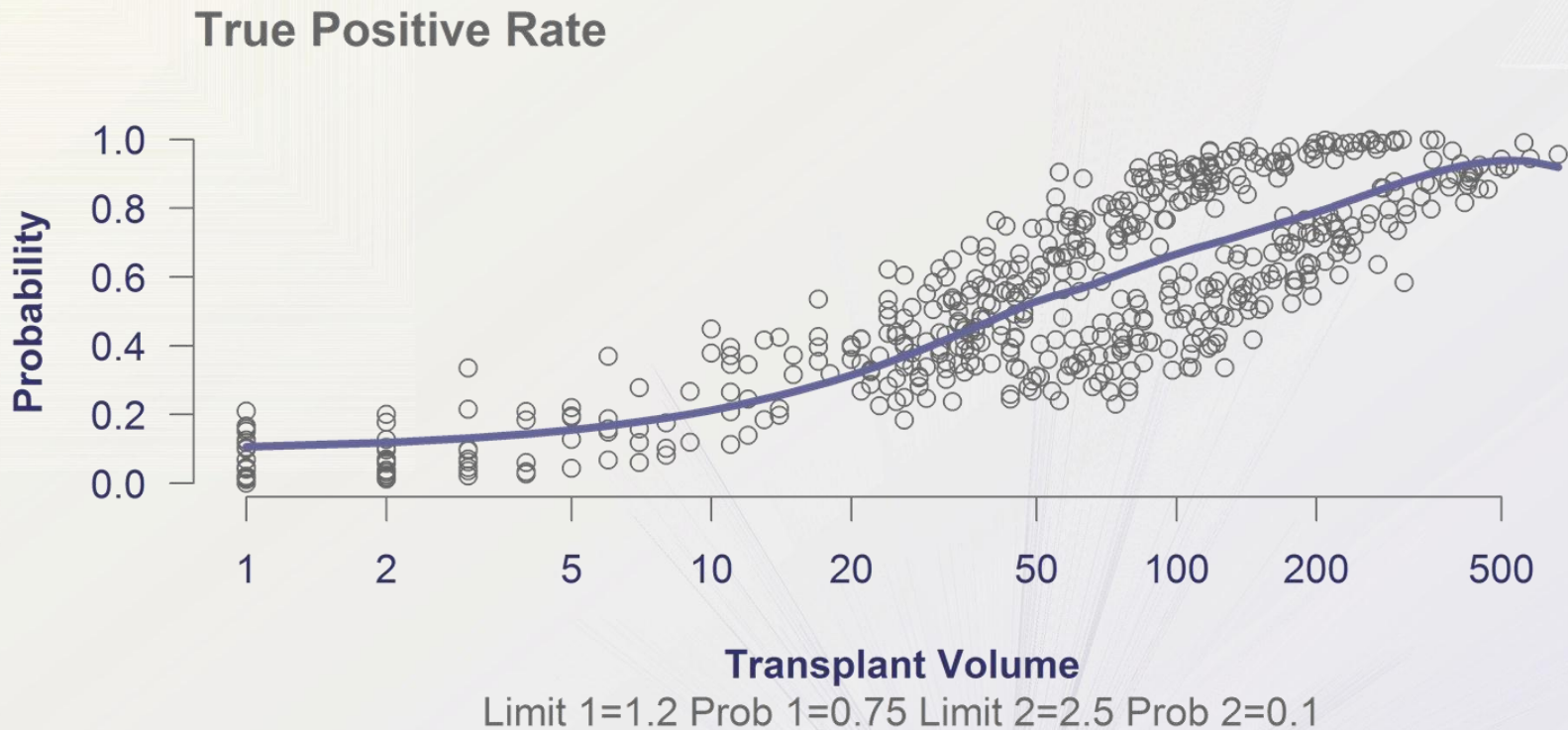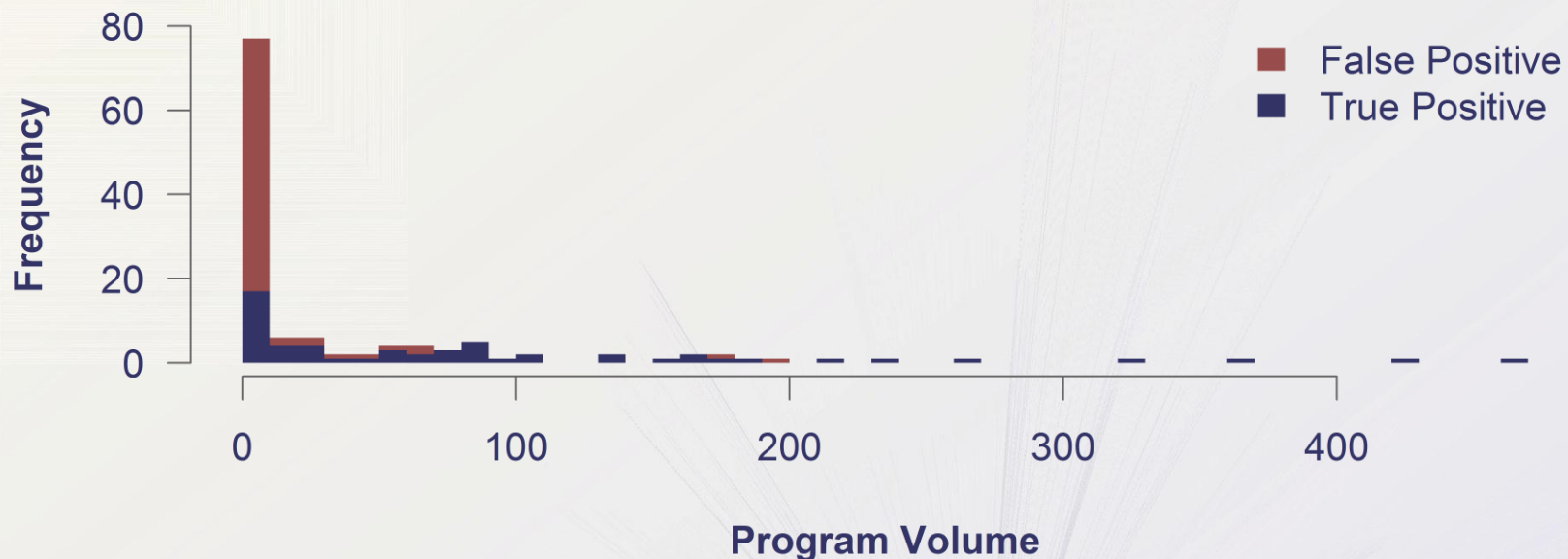
Current Flag Criteria

- For **small-volume programs** (<10 transplants in 30 months): Flag if >0 events.

- For **medium- and large-volume programs** (10 or more transplants in 30 months): Flag if **all 3 criteria** (below) are met.
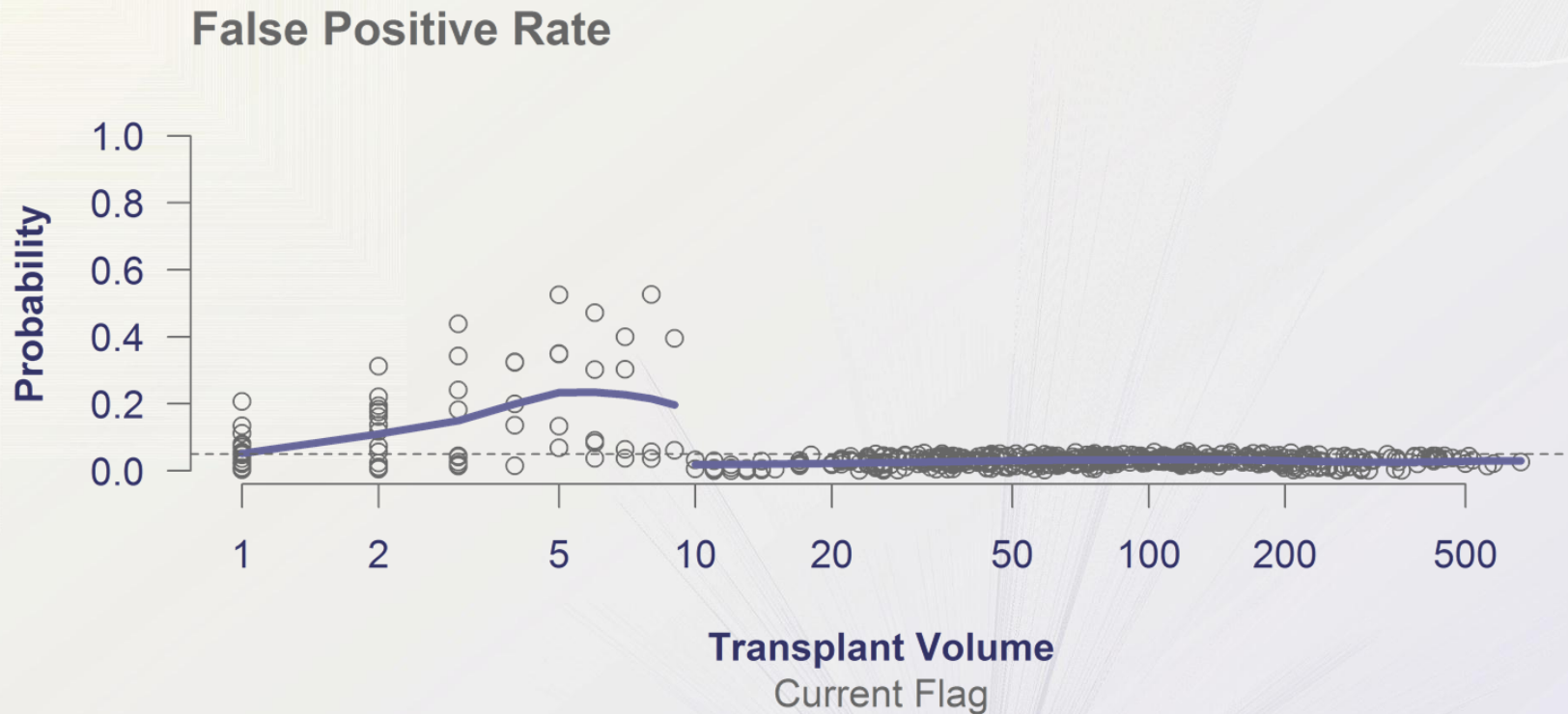
O – E > 3.0

O/E > 1.5

One-sided p-value <0.05
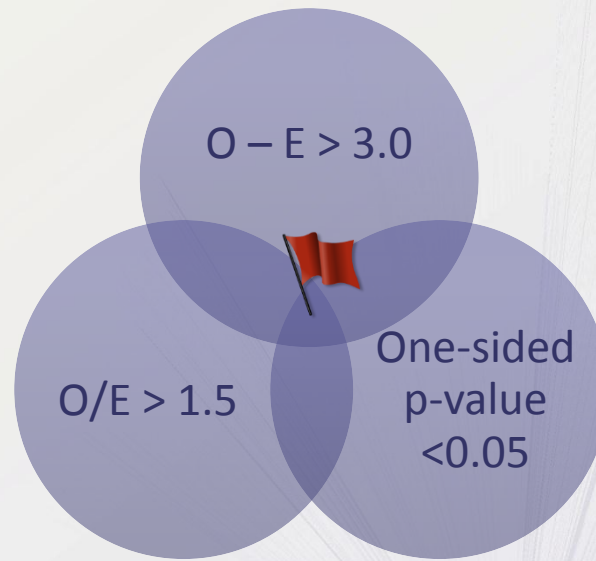
The Current Algorithm Flags a Lot of Small Volume Programs by Mistake

Current Flag False Positive Rate: All

False Positive Rate — Probability vs Transplant Volume (Current Flag)
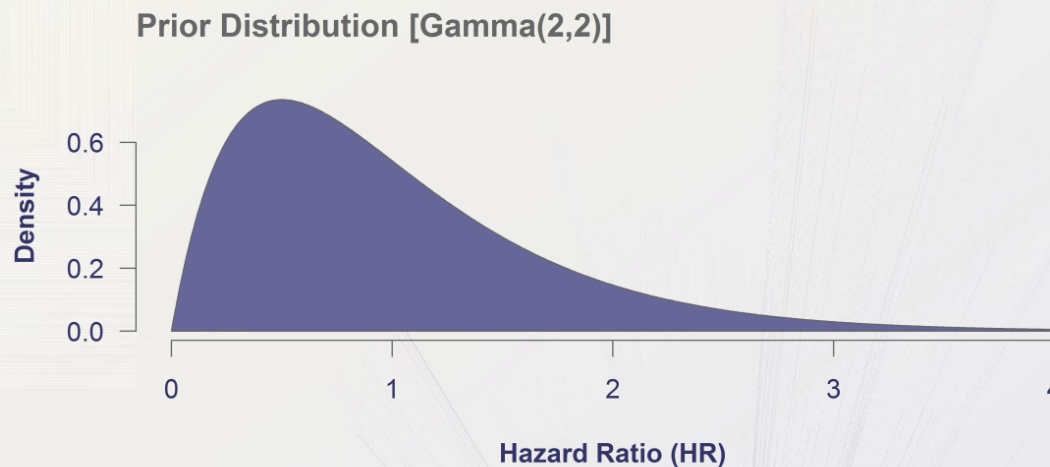
Current Flag True Positive Rate: All
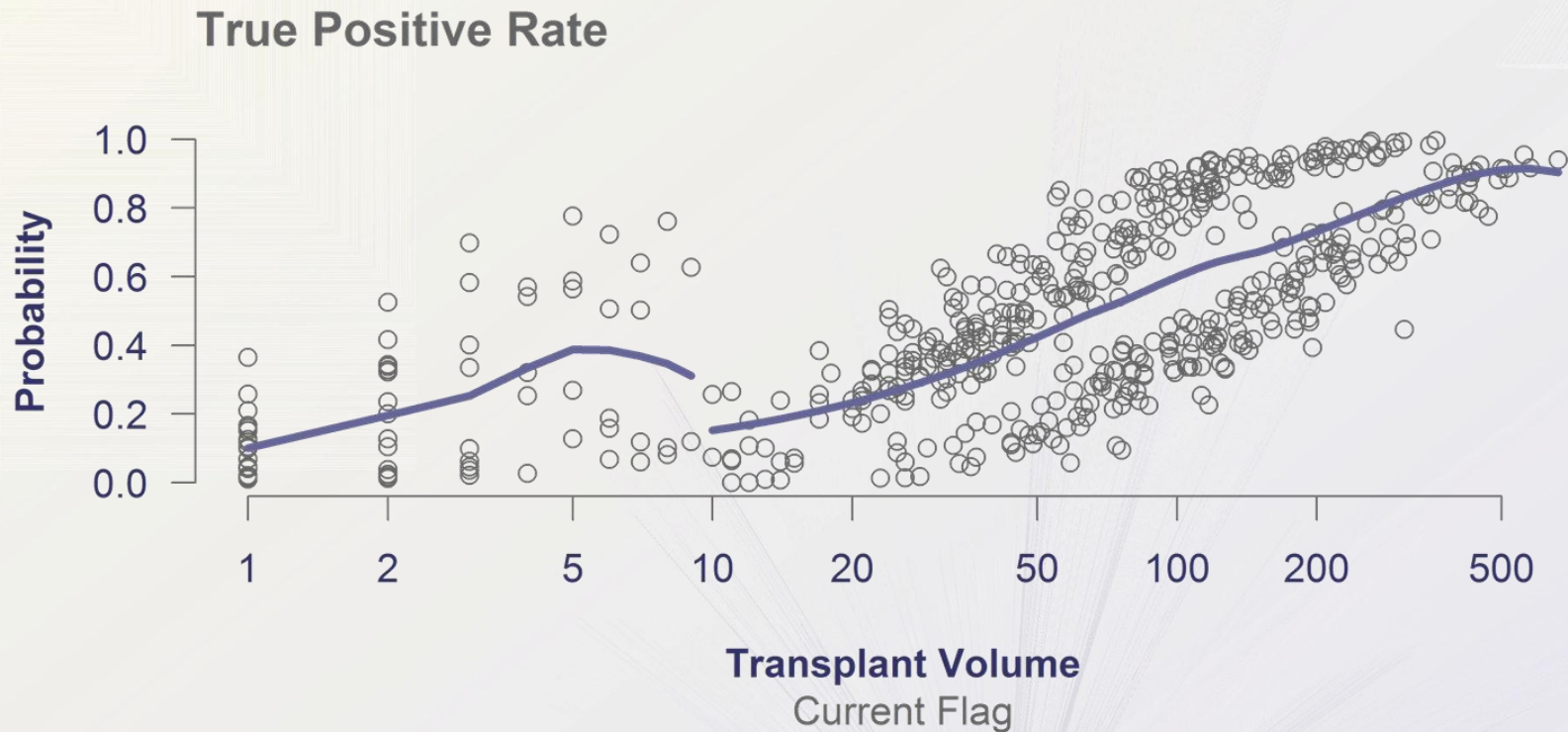
Prior Distribution

SRTR has explored various priors and with the assistance of the SRTR Technical Advisory Committee decided to pursue a relatively weak prior.

Mean 1, Variance 0.5, Gamma (2,2)

Prior Distribution [Gamma(2,2)]

If SRTR used the past performance of programs to determine the prior, it would be so strong that it would overwhelm the data.

Conclusions