

# Biomarkers: the Good, the Bad and the Ambiguous

Enrique F. Schisterman, PhD  
Epidemiology Branch – DESPR- NICHD



# Outline

---

- **Background**
- Limit of Detection Determination
- Pooling Biomarkers
- Limit of Detection
- Conclusions
- Hybrid and Case Only Design

# Background

---

- Biomarker: A specific physical trait used to measure or indicate the effects or progress of a disease or condition
- Newly developed laboratory methods expand the number of biomarkers on a daily basis

# Methodological Constraints

---

- Cost
- Measurement Error
- Causal Link to Disease

# Outline

---

- Background
- **Limit of Detection Determination**
- Pooling Biomarkers
- Limit of Detection
- Conclusions
- Hybrid and Case Only Design

# Reporting of Biomarker Data

ID	Z		ID	Z		ID	Z
1	3.1		1	3.1		1	3.1
2	1.5		2	ND		2	1.1
3	8.4		3	8.4		3	8.4
4	0.8	Report values < threshold as one half the value of the threshold	4	ND	Report values < threshold as 'not detected'	4	1.1
5	5.4		5	5.4		5	5.4
6	3.2		6	3.2		6	3.2
7	2.0		7	ND		7	1.1
8	5.8		8	5.8		8	5.8
9	13.4		9	13.4		9	13.4
10	2.5		10	2.5		10	2.5
11	1.9		11	ND		11	1.1
12	6.1		12	6.1		12	6.1

- Reporting threshold is equal to 2.2

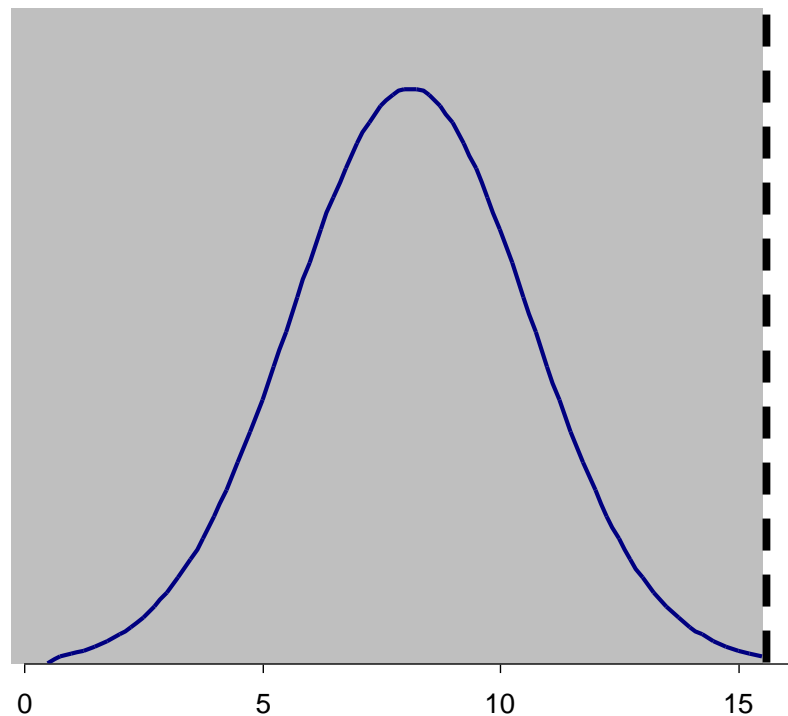
# Conventional Determination of the Limit of Detection (LOD)

## BLANK SERIES

- 10.0
- 5.0
- 8.1
- 7.1
- 4
- 11.3
- 12.0
- 8.0
- 7.7
- 7.0

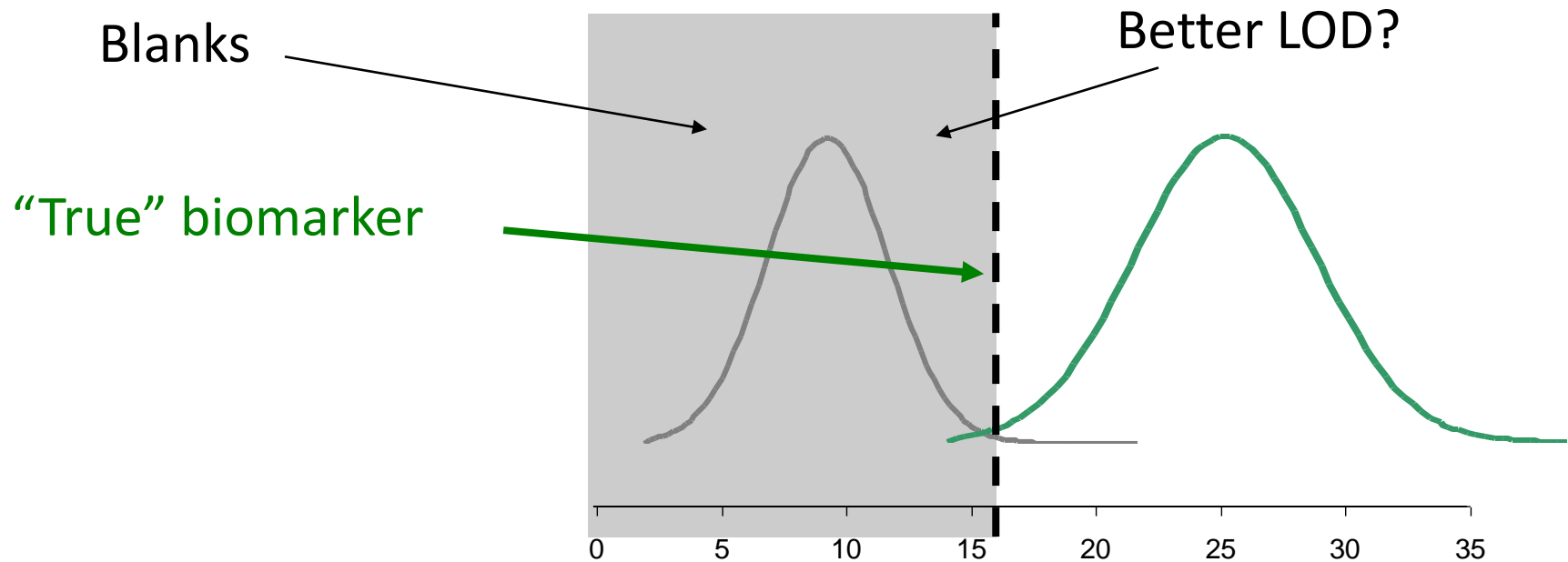
**Mean** = 8.02

**Std Dev** = 2.53



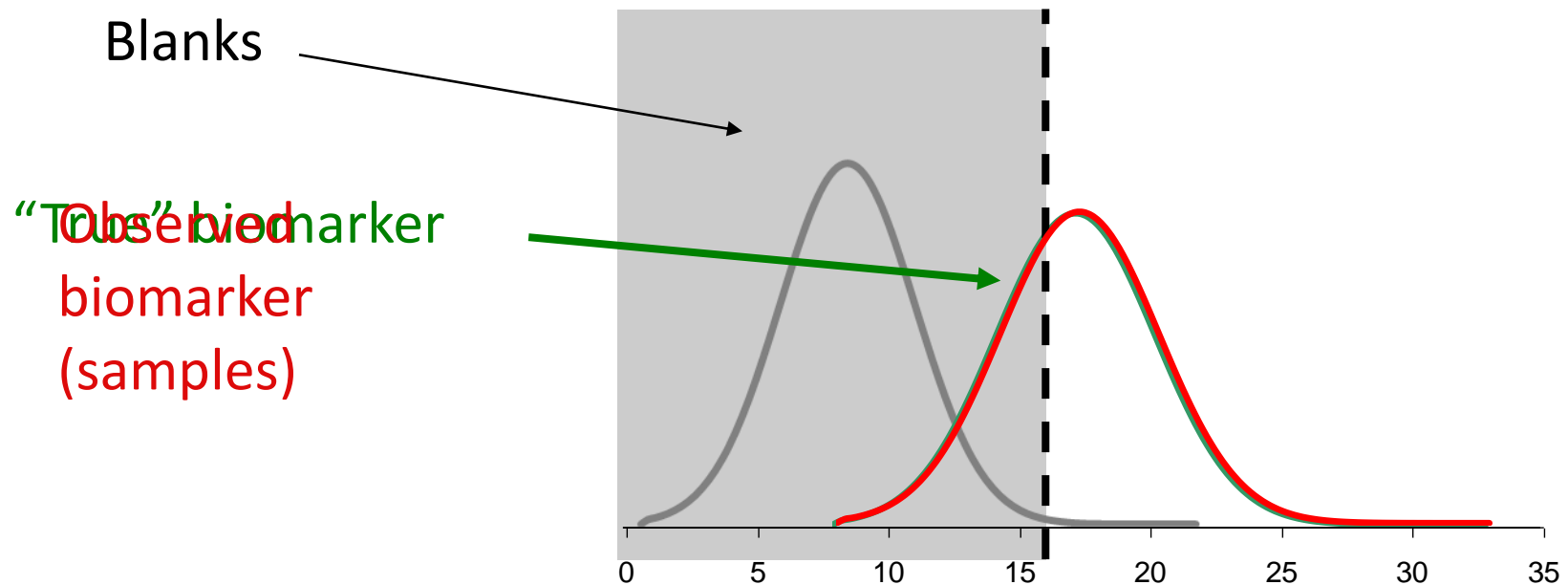
$$LOD = (\mu_{blanks} + 3\sigma_{blanks}) = 15.6$$

# Example of LOD left-censored data



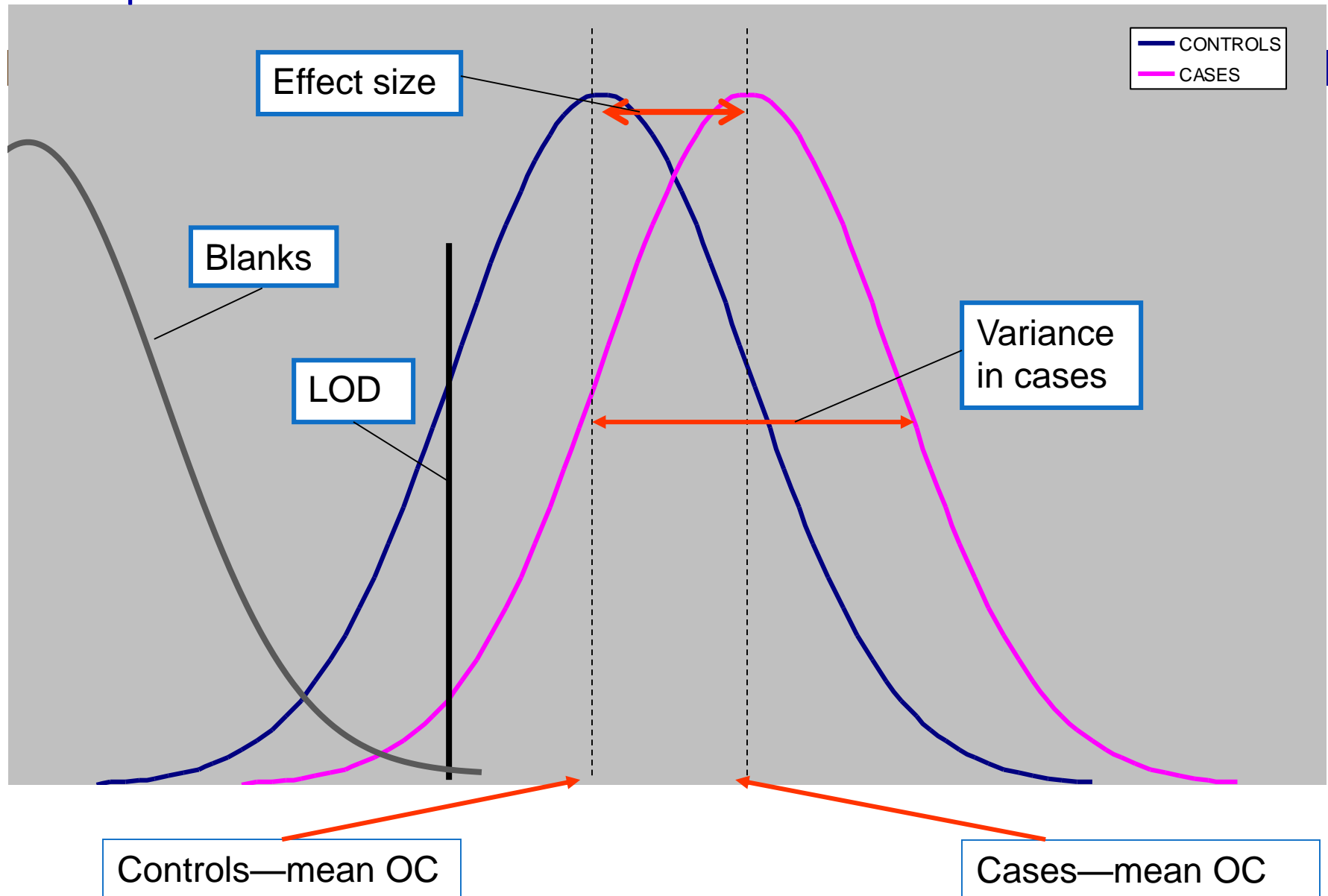


# Example of LOD left-censored data



# Why is this a problem?

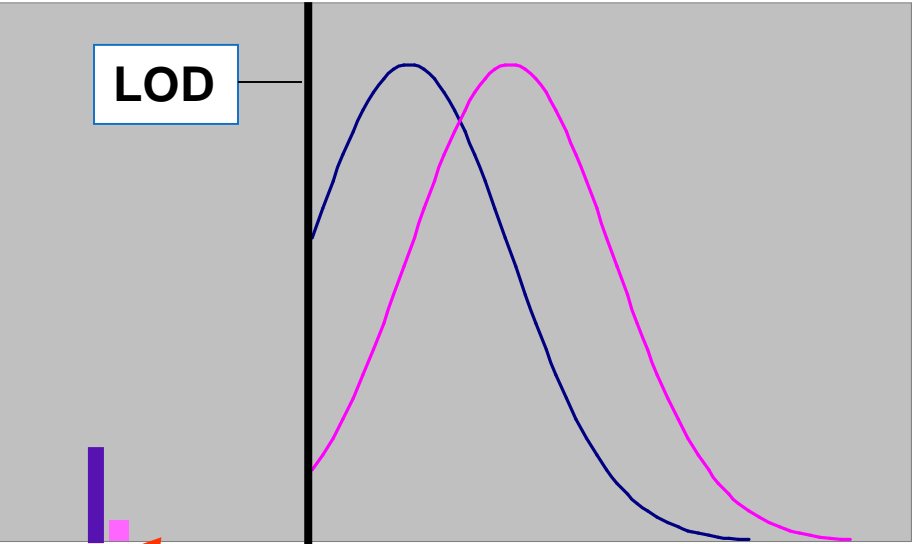
## Comparisons of PCBs in cases and controls



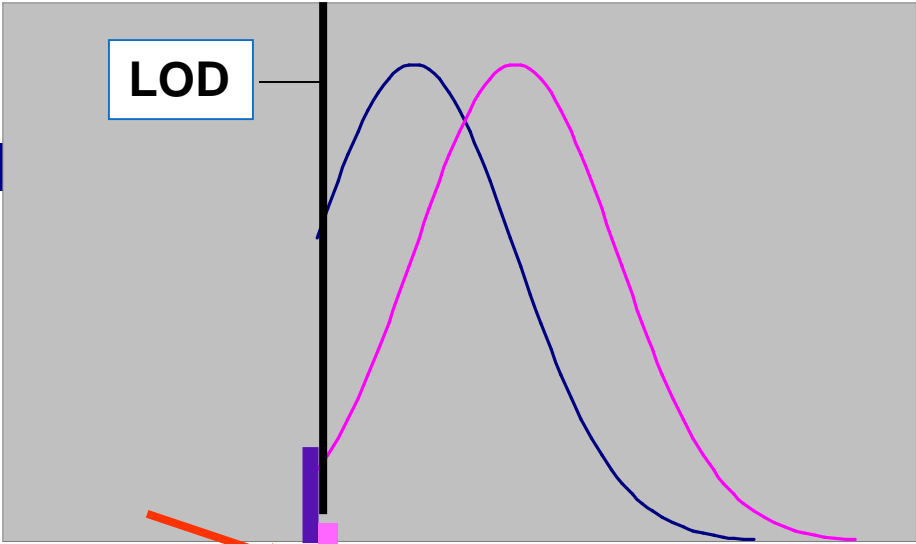
# Approaches for LOD/ missing data

- Simplest approach is substitution
  - Under certain circumstances yield minimal bias
  - Conventionally, values below the LOD are usually
    1. replaced by a. zero; b. the LOD; c.  $\text{LOD}/2$ ; d.  $\text{LOD}/\sqrt{2}$ ;
    2. excluded
    3. retained
- Model based approaches
  - Likelihood models (Perkins et al., AJE 2007)
  - Multiple imputation

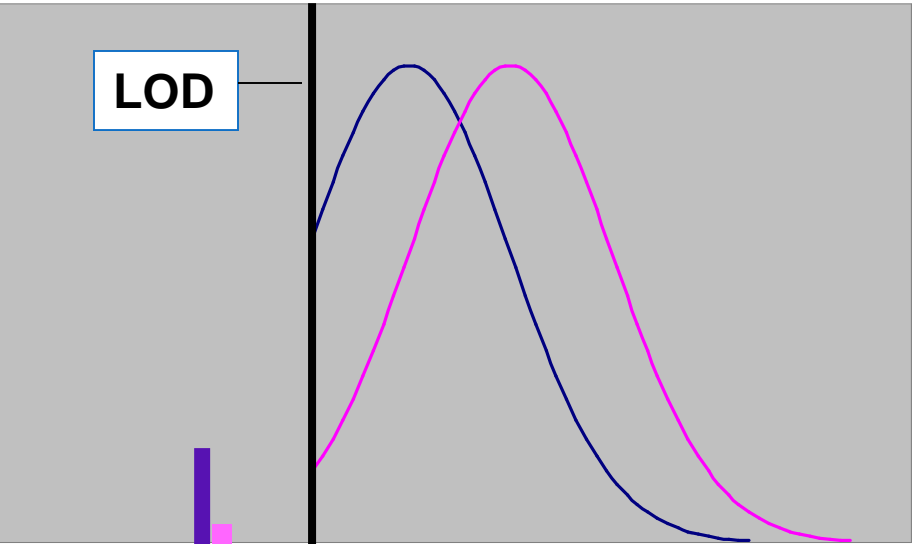
# Imputation and Distribution of Cases and Controls



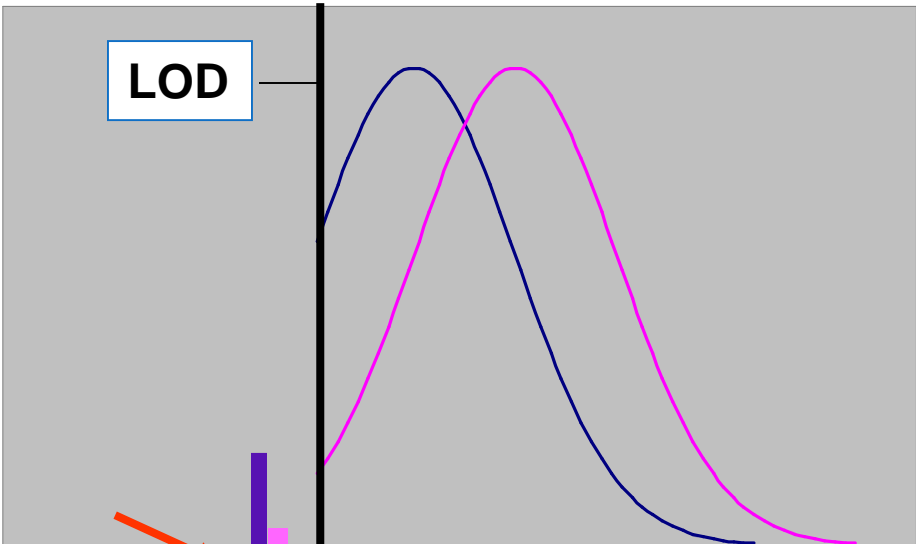
0  
Imputation of zero for values < LOD



2  
Imputation of LOD for values < LOD



1  
Imputation of  $LOD/2$  for values < LOD



1.4  
Imputation of  $LOD/1.4$  for values < LOD

# LOD Simulation

- Purpose: To evaluate the effect of the handling of values below the LOD on risk estimates
- Simulated data from a normal and log normal distribution and varied:
  - Effect size
  - Variance of OCs in the exposure group
  - LOD level
  - Measurement error mean and variance

# Effect of Handling of Values < LOD on %Bias

<i>Method for values &lt; LOD</i>		Parameter values			
		Effect size = .25 SD		Effect size = 2.0 SD	
		LOD high	LOD low	LOD high	LOD low
<b>1. Replace by</b>	<b>a. Zero</b>	-59.0	-25.1	-118.8	-114.2
	<b>b. LOD</b>	-187.1	-40.8	-365.2	-360.7
	<b>c. LOD/2</b>	-71.3	-18.1	-82.3	-118.5
	<b>d. LOD/<math>\sqrt{2}</math></b>	-79.7	-15.9	-65.1	-119.2
<b>2. Exclude (truncated)</b>		-314.2	-265.3	-211.9	-248.0
<b>3. Retain (observed)</b>		-11.5	-11.7	-12.10	-11.98

\*LOD “low” indicates 1.6 SDs below the mean of controls, resulting in imputed values for a small number of data points. LOD “high” indicates 1 SD above the mean of the controls, resulting in imputed values for a large number of both controls and cases

# LOD—Conclusions

- Choice of how to handle values below the LOD can result in a loss of accuracy in estimating risk
- Retaining observed values below the LOD produces the least biased estimates
- Substitution of  $\text{LOD}/\sqrt{2}$  for values below the LOD produces not terribly biased estimates

# Outline

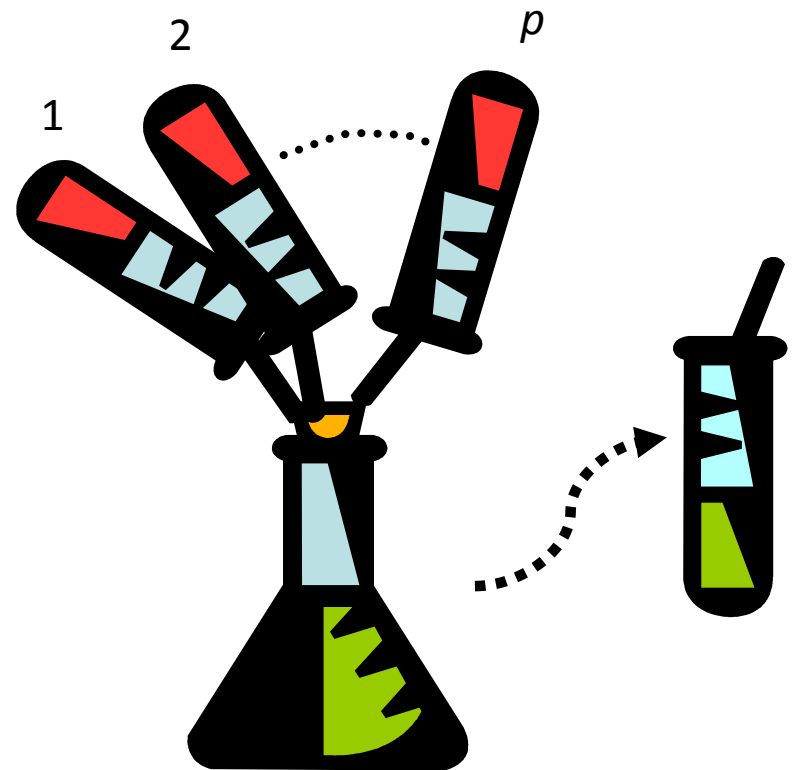
---

- Background
- Limit of Detection Determination
- **Pooling Biomarkers**
- Limit of Detection
- Conclusions
- Hybrid and Case Only Design

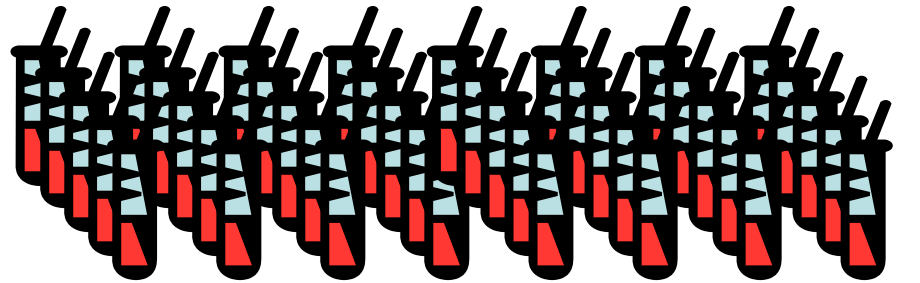


# What is pooling?

- Physically combining several individual specimens to create a single mixed sample
- Pooled samples are the average of the individual specimens

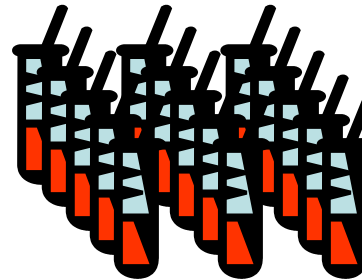


# Random Sample of Biospecimens



FULL DATA

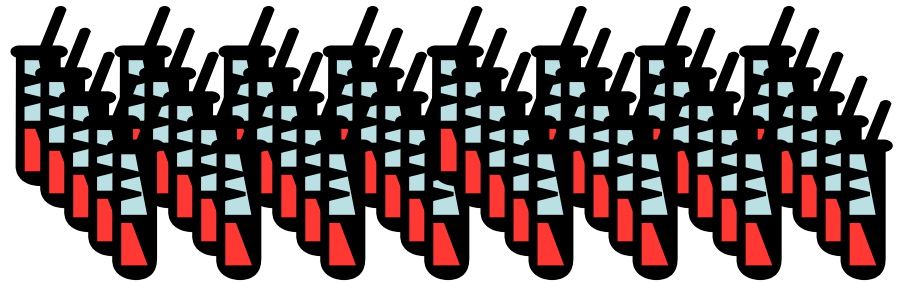
N = 40 Individual  
Biospecimens



RANDOM SAMPLE:

Randomly select 20 samples

# Pooling Biospecimens



FULL DATA

N = 40 Individual  
Biospecimens



POOLED DATA:

40 samples in groups of 2

# LOD and Pooling

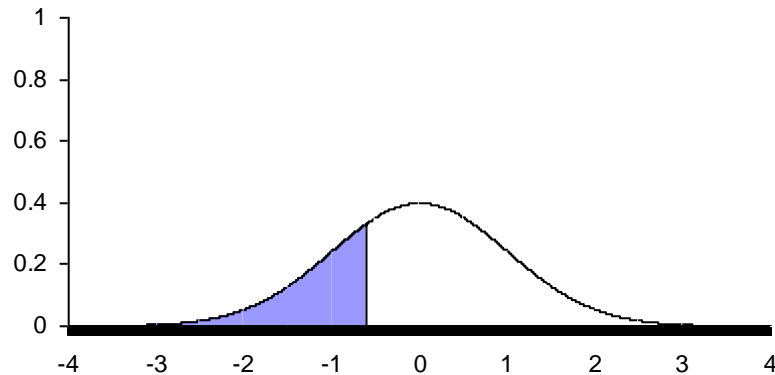
---

Unpooled Specimens

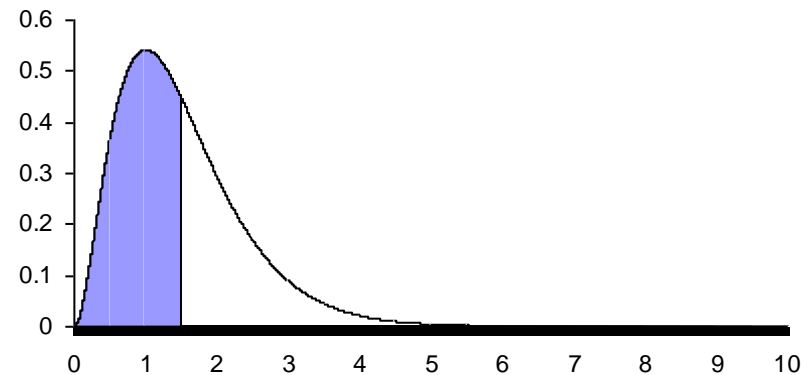
Pooled Specimens

# Effect of Pooling on Markers Affected by an LOD

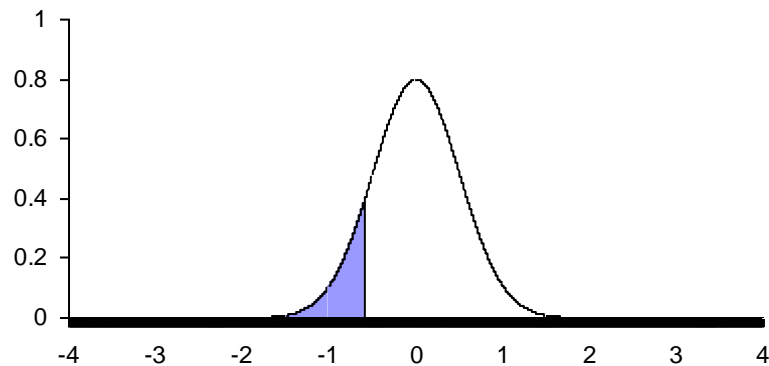
Un-pooled



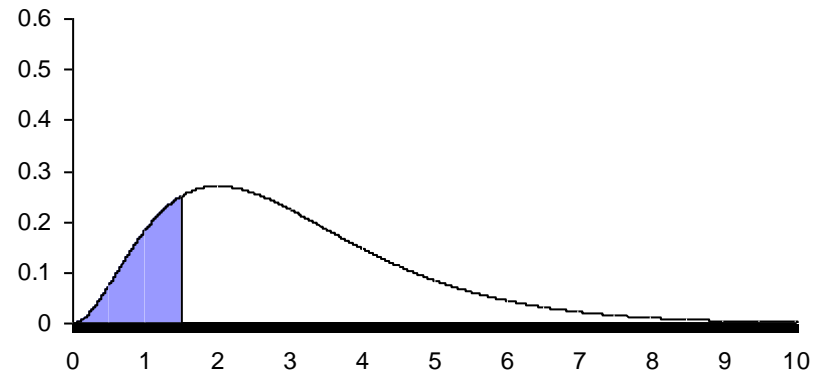
Un-Pooled



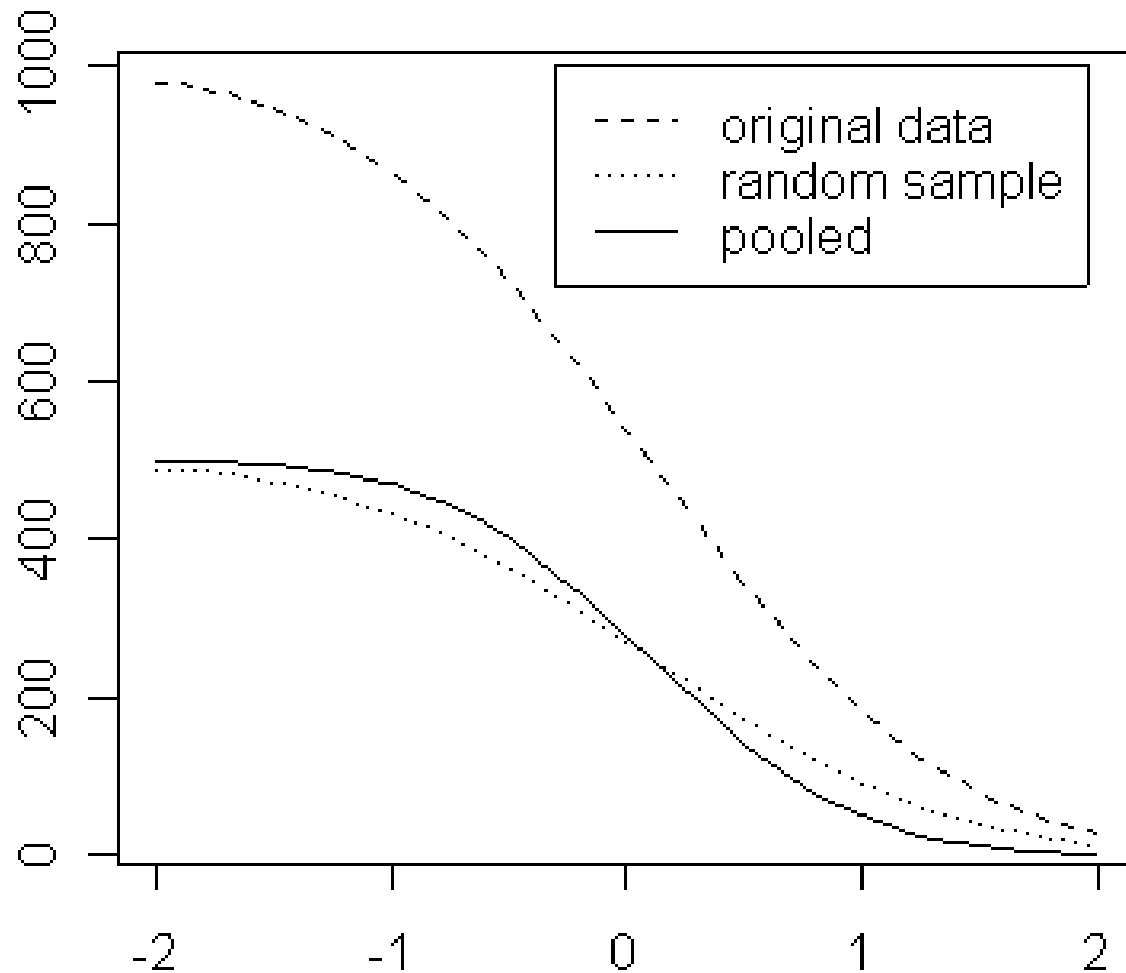
Pooled



Pooled

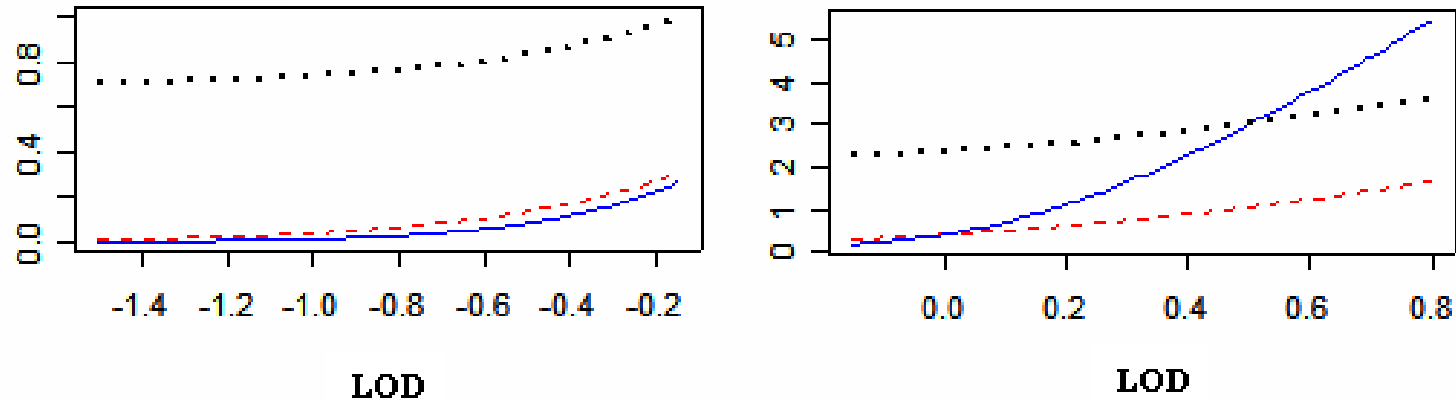


# Comparison of the Number of Observations Above the LOD for Standard Normal Data

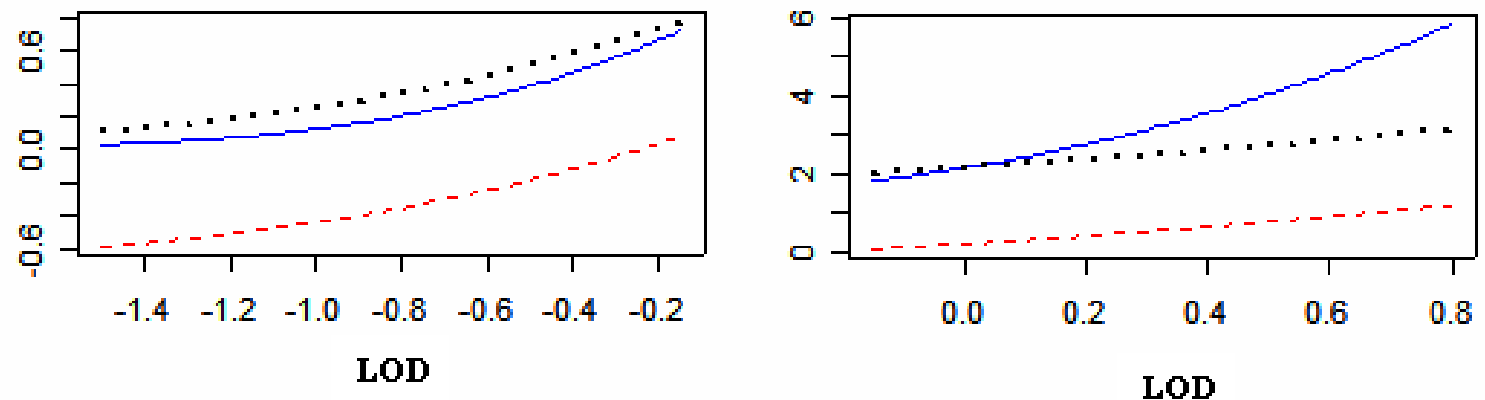


# Efficiency of the Mean and Variance

## Variance of Estimated Mean



## Variance of Estimated Variance



--- FULL DATA      — POOLED      ... RANDOM

# Pooling and Random Sampling

- Pooling advantages
  - Reduces the number of assays we need to test
  - Estimates the mean extremely efficiently
  - Cost-effective
- Random sampling advantages
  - Reduces the number of assays we need to test
  - Very easy
  - Variance is estimated very well



# Hybrid Design: Pooled—Unpooled

- Create a sample of both pooled and unpooled samples that takes advantage of the strengths of both the pooling and random sampling designs
  - Reduces number of tests to perform
  - Cuts overall costs
  - Gains efficiency (by using pooling technique)
  - Accounts for different types of measurement error without replications
    - Pooling error
    - Random measurement error
    - LOD

# Setup of Hybrid Design

Hybrid Sample S:  $X_1, \dots, X_5, Z_1, \dots, Z_{15}$

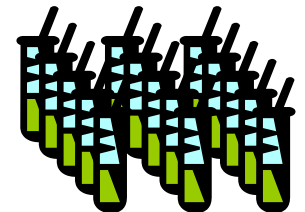
Unpooled:

$X_1, \dots, X_5$



Pooled:

$Z_1, \dots, Z_{15}$

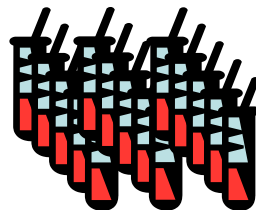


In General

Hybrid Sample S:  $X_1, \dots, X_{[\alpha n]}, Z_1, \dots, Z_{[(1-\alpha)n]}$

Unpooled:

$X_1, \dots, X_{[\alpha n]}$



Pooled:

$Z_1, \dots, Z_{[(1-\alpha)n]}$



$\alpha$  is the proportion of unpooled samples

# Maximum Likelihood Estimators



In order to estimate the  
variance,  $\alpha$  cannot be zero.

# Hybrid Design Example: IL-6

- Measured IL-6 on 40 MI cases and 40 controls
- Biological specimens were randomly pooled in groups of 2, for the cases and controls separately, and remeasured
- We want to evaluate the discriminating ability of this biomarker in terms of AUC

# Hybrid Design Example: IL-6

	n	$\alpha_x$	$\alpha_y$	AUC	Var(AUC)
Empirical	40	1.00	1.00	0.640	0.0036
Hybrid design: Optimal $\alpha$	20	0.40	0.35	0.621	<b>0.0049</b>
Random sample: $\alpha=1$	20	1.00	1.00	0.641	0.0071

Hybrid design reduced the variability of Var(AUC) by 32% as compared to taking only a random sample

# Summary—Hybrid Design

- Hybrid design is a more efficient way to estimate the mean and variance of a population
- Cost-effective
- Yields estimate of measurement error without requiring repeated measurements
- Here we focus on normally distributed data, but can be applied to other distributions as well

# Outline

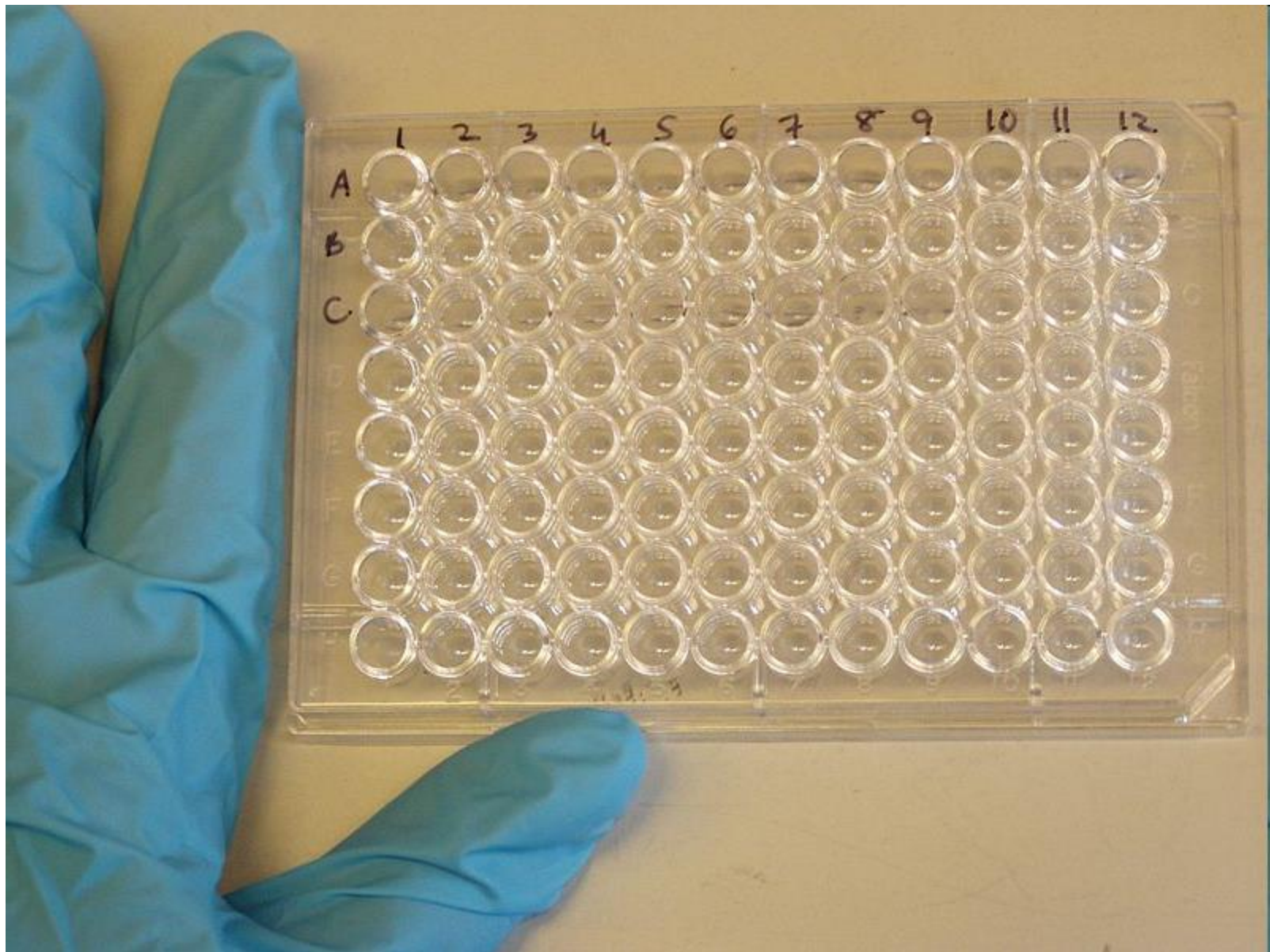
---

- Background
- Limit of Detection Determination
- **Pooling Biomarkers**
- Limit of Detection
- Conclusions
- Hybrid and Case Only Design

## Measurement of GCSF

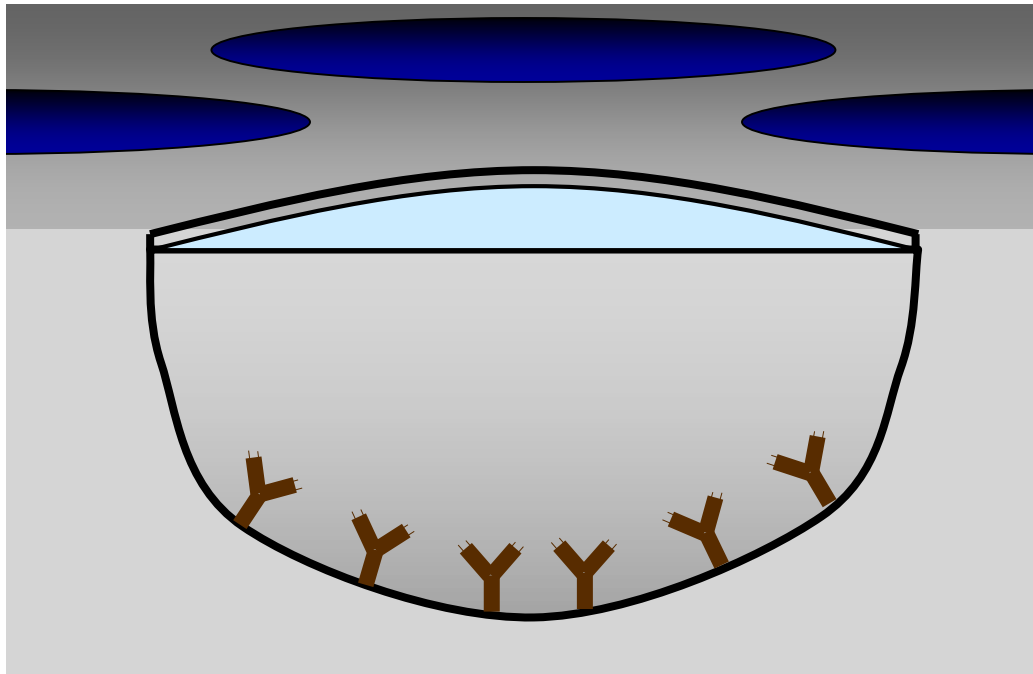
- Chemiluminescence assays
  - A 96-well plate
  - Antibody against the biomarker of interest
  - A set of standards of known biomarker concentration included in each batch
  - A set of unknowns for which we would like to know biomarker concentrations
  - A light emitting molecule that binds to bound biomarker





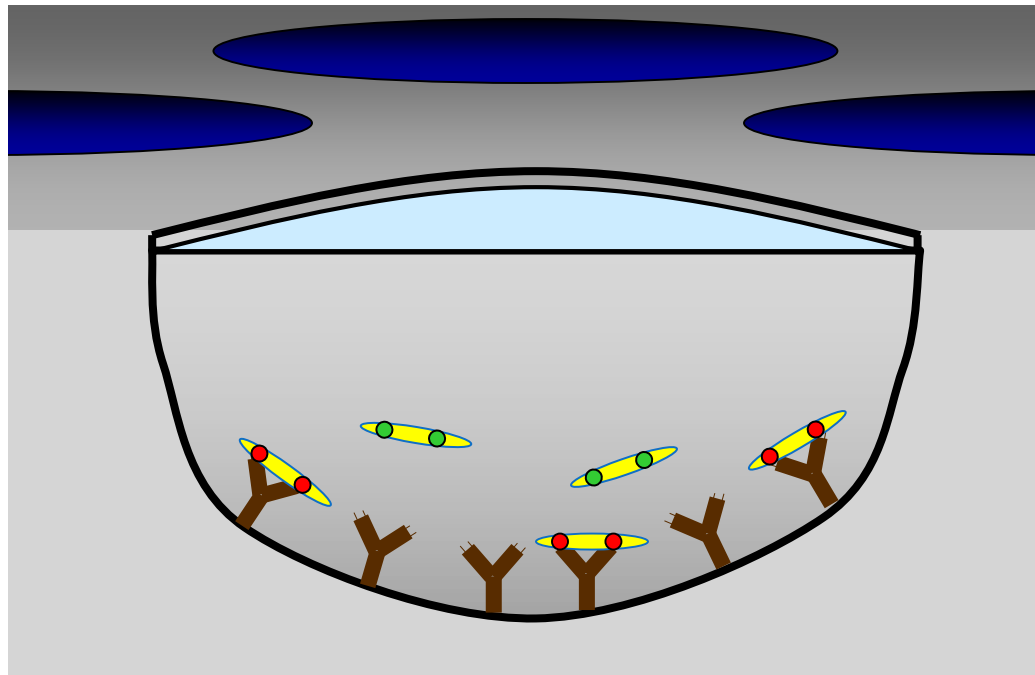
# Measurement of cytokines by chemiluminescence assay

- Cytokines are not measured directly
  - Antibodies against analyte(s) coat wells



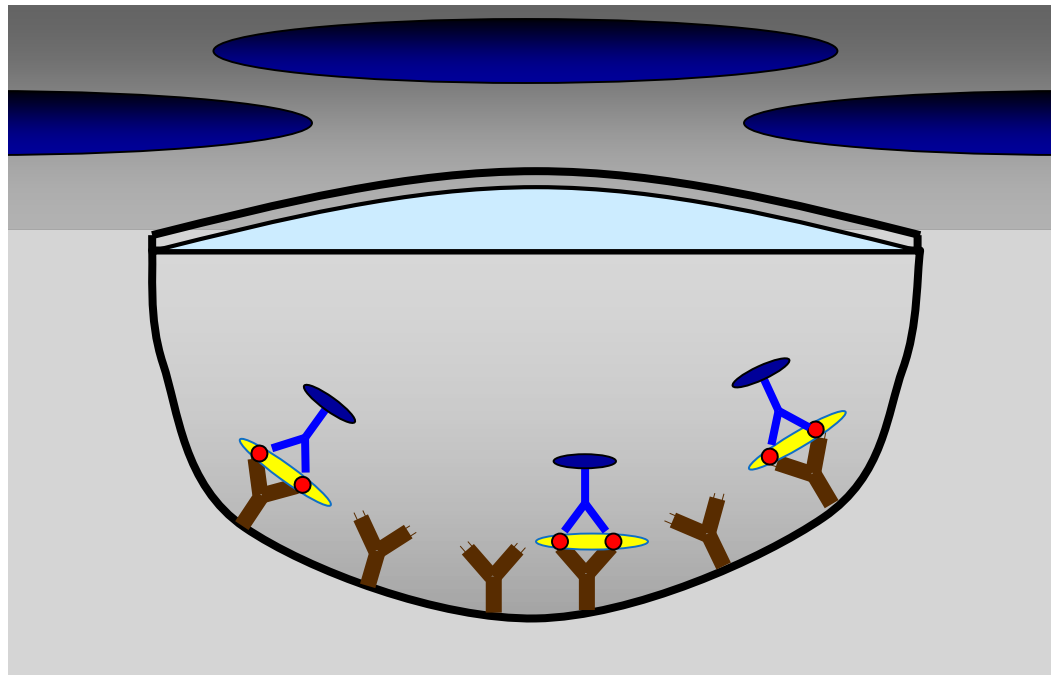
# Measurement of cytokines by chemiluminescence assay

- Samples added, analyte binds to antibodies
- Unbound proteins is washed away



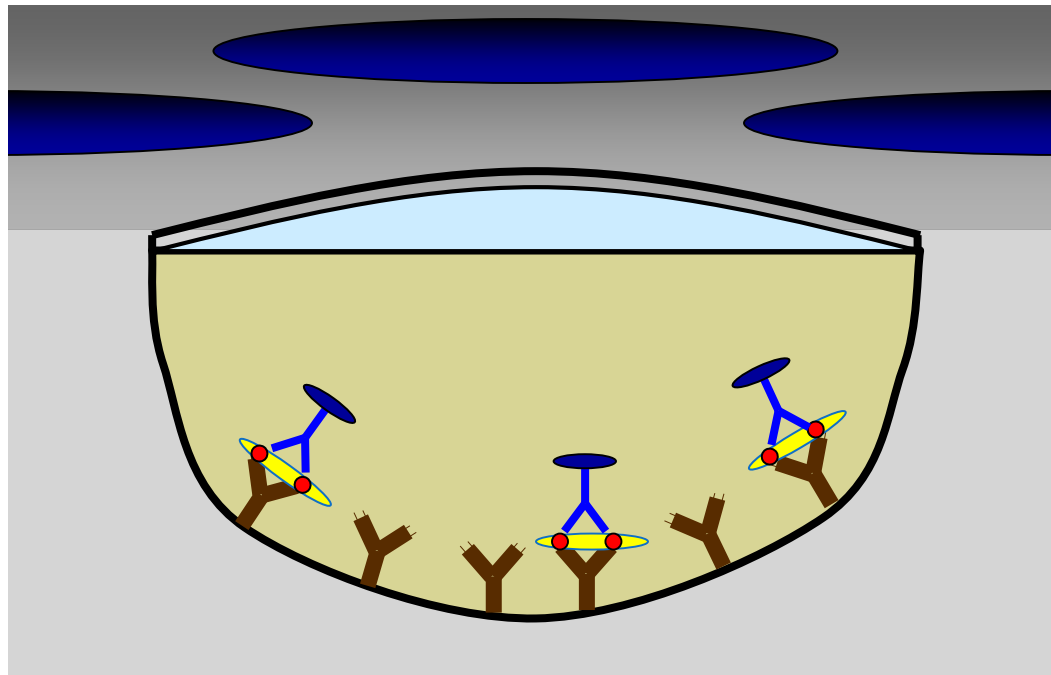
# Measurement of cytokines by chemiluminescence assay

- A 'tag' is added to the assay that binds to the protein – antibody complex that produces color
- The intensity of the color is measured



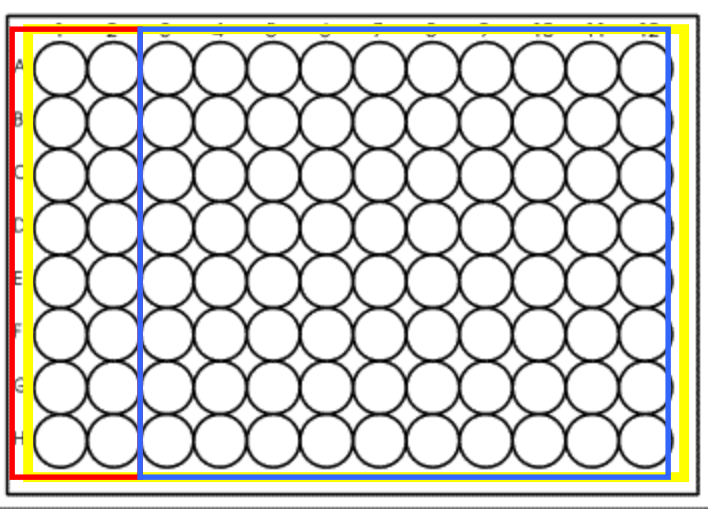
# Measurement of cytokines by chemiluminescence assay

- A 'tag' is added to the assay that binds to the protein – antibody complex that produces color
- The intensity of the color is measured



# ELISA/multiplex layout

- Step 1: prepare antibodies mixture and add to plate
- Step 2: prepare calibrators, add to plate
- Step 3: prepare unknowns, add to plate



# Use of chemiluminescence assays for measuring protein concentrations

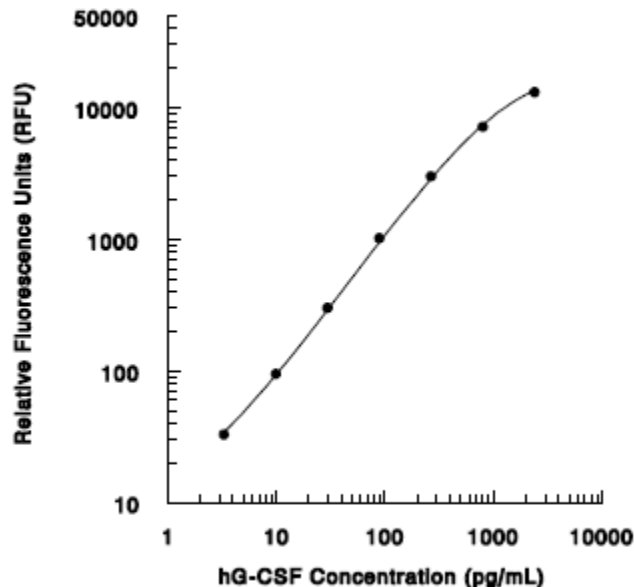
- Use calibration to convert relative measures to the desired unit of concentration
  - from **optical density** in relative fluorescence units (RFU) to **concentration** in pg/mL
- Current practice is per assay calibration
  - Results in potentially large calibration datasets used only minimally in current practice

# Calibrating the assay

## The standard curve

### *Typical Data*

This human G-CSF standard curve is provided only for demonstration. A standard curve must be generated each time an assay is run, utilizing values from the Standard Value Card included in the Base Kit.



Standard	pg/mL	RFU	Average	Corrected
Blank	0	34 37	36	—
1	2400	12921 13409	13165	13129
2	800	7125 7308	7217	7181
3	267	3042 3044	3043	3007
4	89	1008 1103	1056	1020
5	30	322 353	338	302
6	10	127 136	132	96
7	3	67 72	70	34



# Calibrating the assay

## The standard curve

This human G-CSF standard curve is provided only for demonstration. **A standard curve must be generated each time an assay is run**, utilizing values from the Standard Value Card included in the Base Kit.

- Potential variation in the relation between relative fluorescence and concentration
  - Chromophore potentially affected by temperature, humidity, etc

# GCSF and miscarriage in the CPP

- Case-control study nested in the Collaborative Perinatal Project study cohort
  - 462 miscarriage case observations
  - 482 non-miscarriage control observations
- Serum biospecimens from early pregnancy, prior to miscarriage onset
- For  $n = 944$ , 24 assays were used

# Circulating levels of cytokines during pregnancy: thrombopoietin is elevated in miscarriage

Brian W. Whitcomb, Ph.D.,<sup>a,b</sup> Enrique F. Schisterman, Ph.D.,<sup>a</sup> Mark A. Klebanoff, M.D.,<sup>a</sup> Mona Baumgarten, Ph.D.,<sup>b</sup> Xiaoping Luo, M.D.,<sup>c</sup> and Nasser Chegini, Ph.D.<sup>c</sup>

**TABLE 2**

Crude and adjusted odds ratio estimates from conditional logistic regression models of risk of miscarriage.

	Unadjusted models		Adjusted model <sup>a</sup>	
	OR	[95% CI]	OR	[95% CI]
Th1-type cytokines <sup>b</sup>				
IL-1 $\beta$	0.90	[0.77, 1.05]	0.76	[0.59, 0.97]
IFN- $\gamma$	1.01	[0.87, 1.18]	1.09	[0.91, 1.30]
TNF- $\alpha$	0.92	[0.79, 1.08]	0.93	[0.77, 1.13]
Th2-type cytokines <sup>b</sup>				
IL-1ra	1.01	[0.89, 1.14]	1.06	[0.90, 1.24]
IL-4	1.00	[0.86, 1.17]	1.05	[0.88, 1.25]
IL-6	0.99	[0.86, 1.15]	1.19	[0.96, 1.48]
Growth factors <sup>b</sup>				
G-CSF	0.84	[0.72, 0.99]	0.78	[0.64, 0.95]
TPO	1.09	[0.95, 1.25]	1.16	[1.00, 1.36]

Note: OR = odds ratio; CI = confidence interval; IL = interleukin; RA = receptor antagonist; G-CSF = granulocyte colony stimulating factor; IFN = interferon; TNF = tumor necrosis factor; TPO = thrombopoietin.

<sup>a</sup> Adjusted model included terms for all cytokines, maternal age, and gestational age at sample collection.

<sup>b</sup> Biomarkers were standardized by dividing assay-determined concentration by the standard deviation among controls.

# Circulating levels of cytokines during pregnancy: thrombopoietin is elevated in miscarriage

*Brian W. Whitcomb, Ph.D.,<sup>a,b</sup> Enrique F. Schisterman, Ph.D.,<sup>a</sup> Mark A. Klebanoff, M.D.,<sup>a</sup> Mona Baumgarten, Ph.D.,<sup>b</sup> Xiaoping Luo, M.D.,<sup>c</sup> and Nasser Chegini, Ph.D.<sup>c</sup>*

	Unadjusted model		Adjusted model	
	OR	[95% CI]	OR	[95% CI]
Factor				
GCSF	0.84	[0.72, 0.99]	0.78	[0.64, 0.95]

This estimate is based on the conventional batch specific approach

# Objective

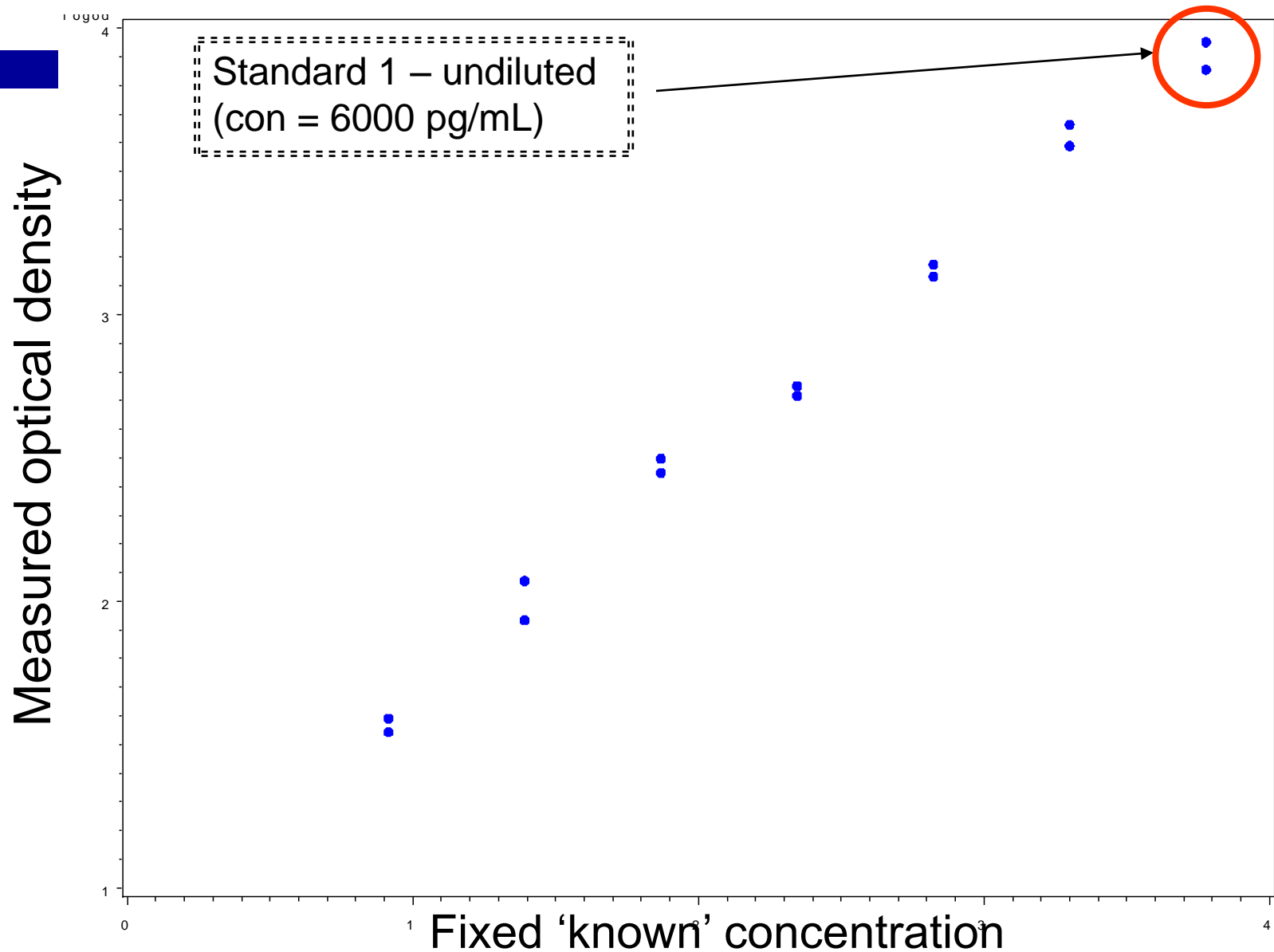
---

- Question: Is the current practice of standard batch-specific approach to calibration the best use of information?
- To evaluate the effect of different approaches for calibration models on models of risk
- To assess bias associated with different approaches

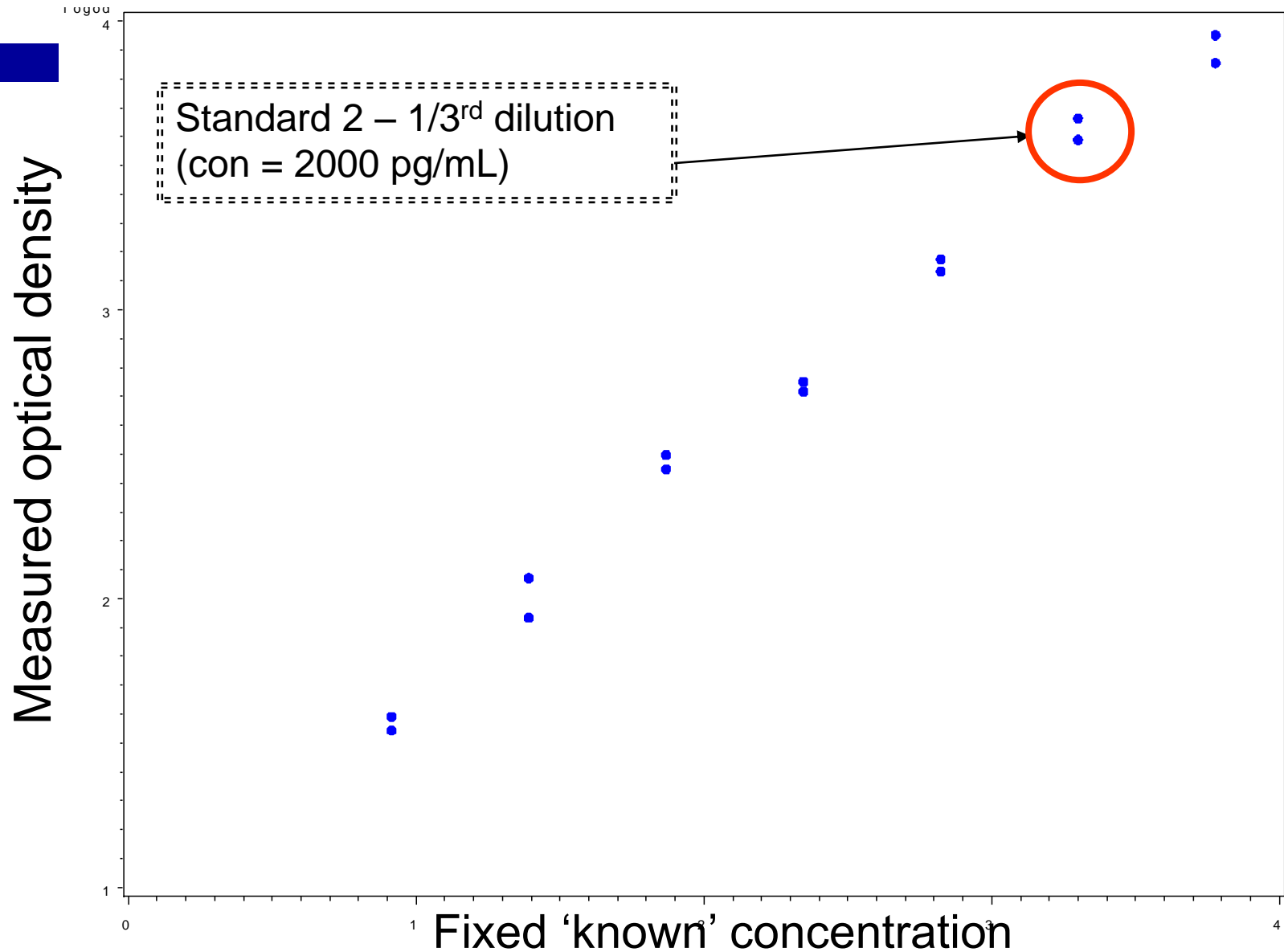
# Data from the calibration experiments

- 24 batches, each with 7 known concentrations measured in replicate
  - Batches varied by
    - Shape
    - Location
    - Agreement between replicates
    - Presence of outliers

# Batch 1 calibration curve - GCSF

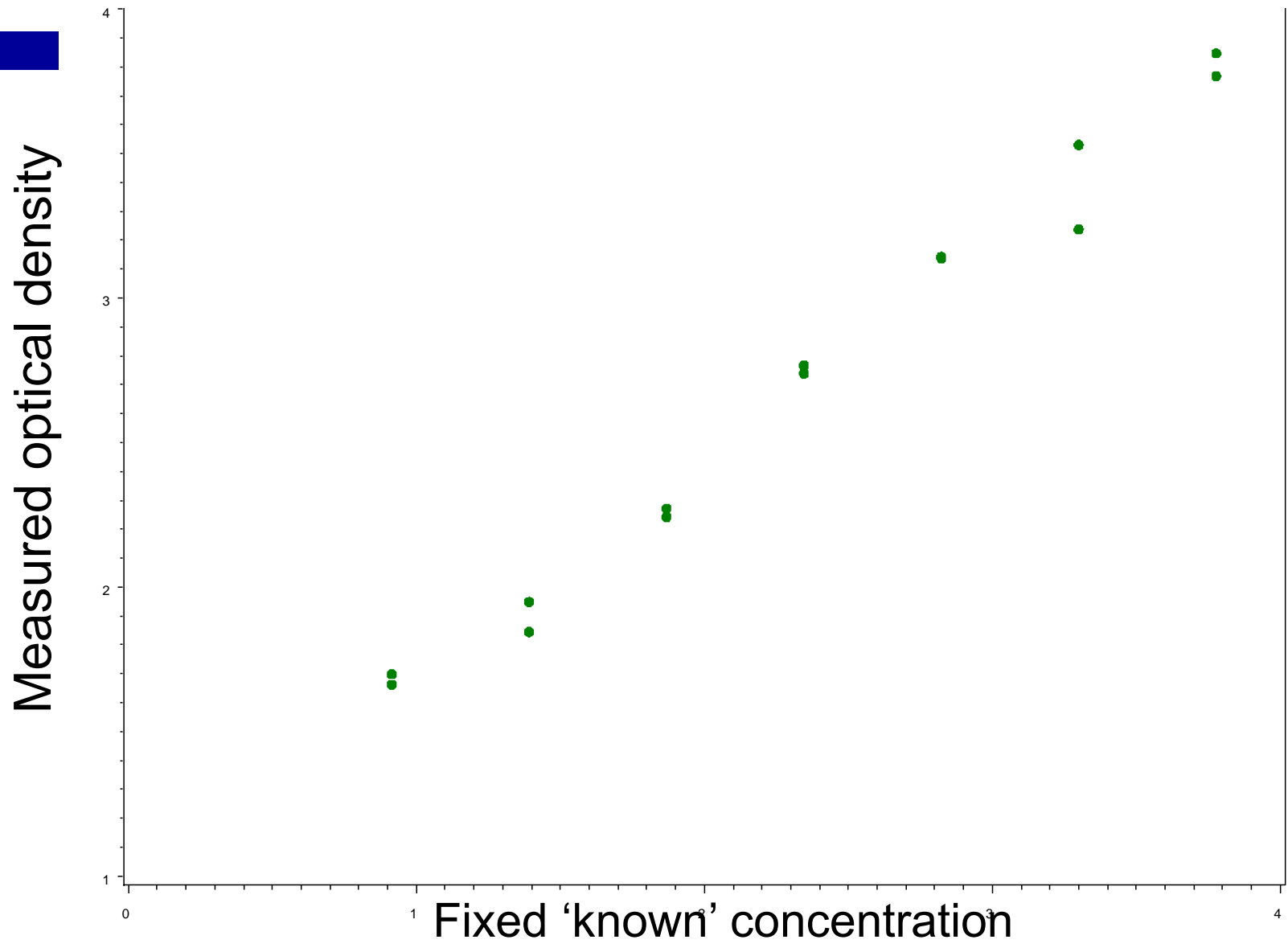


# Batch 1 calibration curve - GCSF

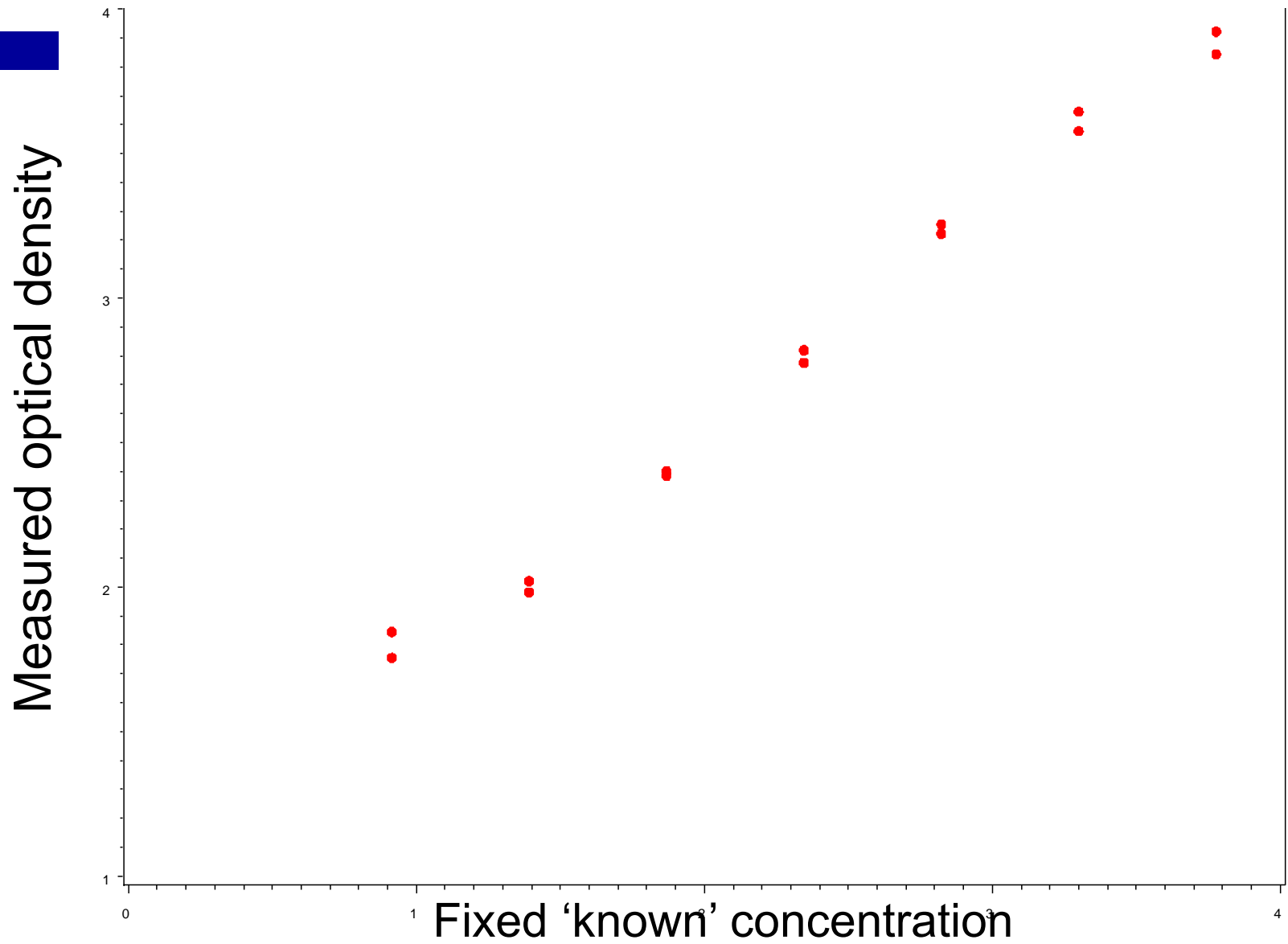




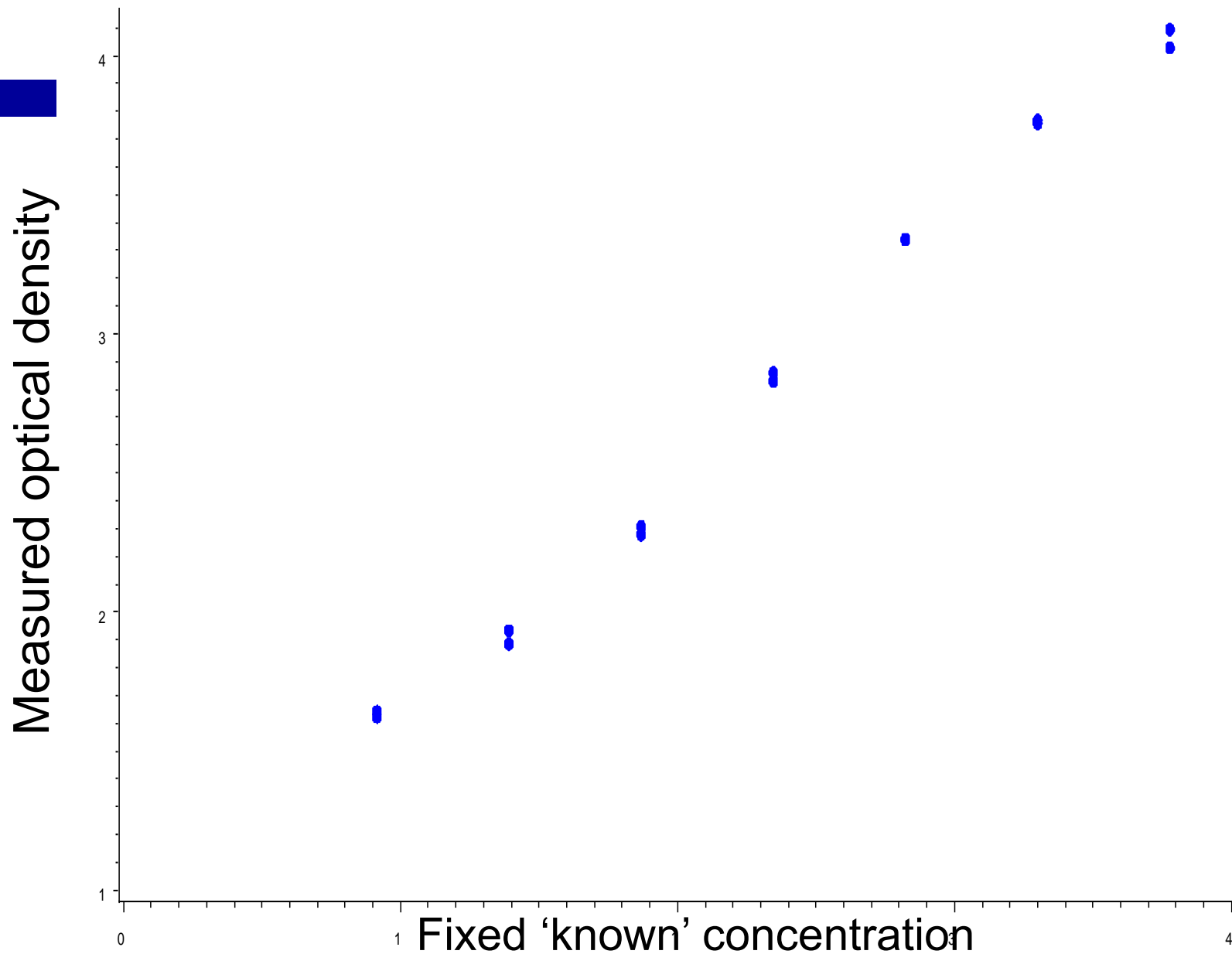
# Batch 2 calibration curve - GCSF



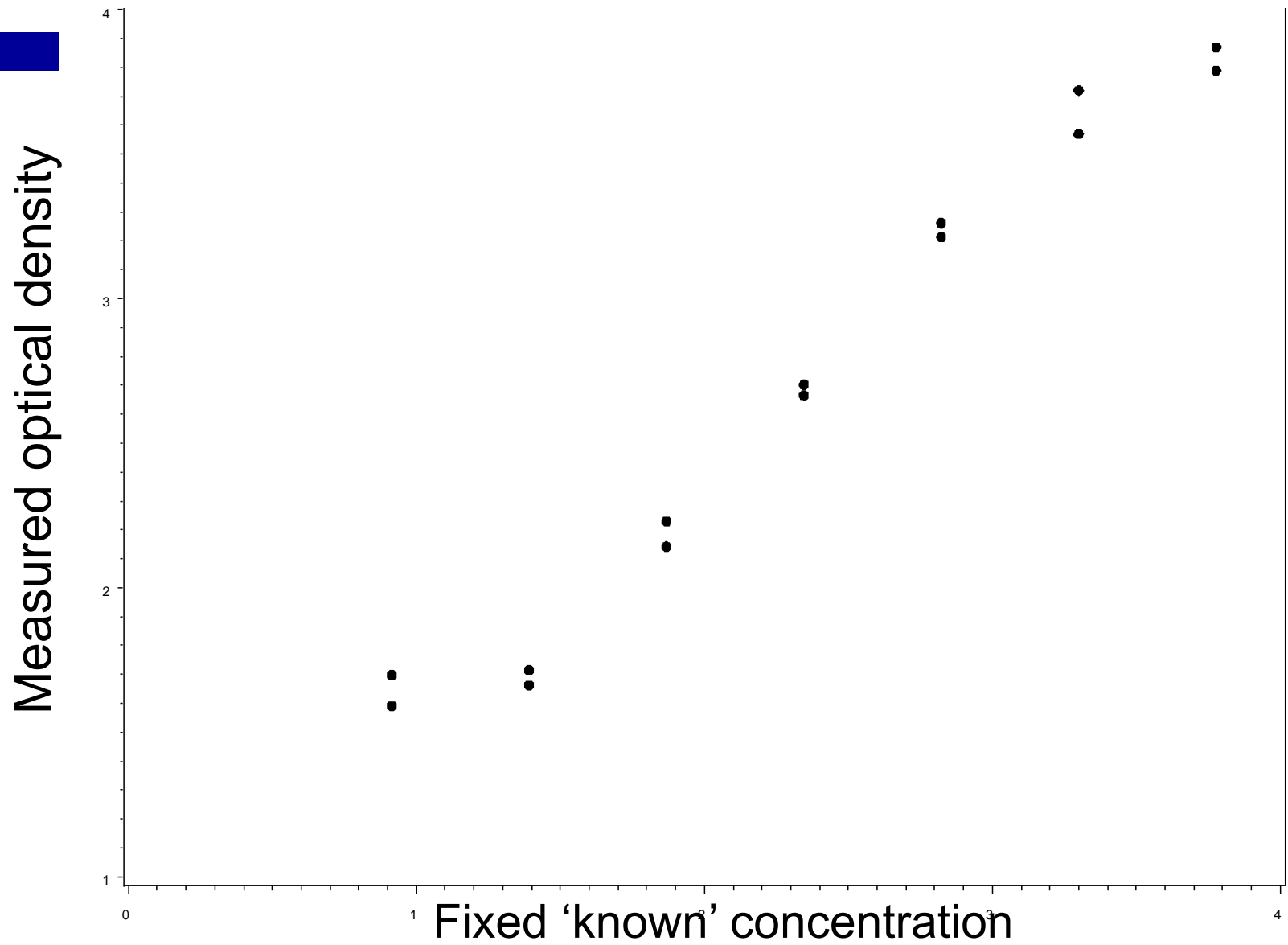
# Batch 3 calibration curve - GCSF



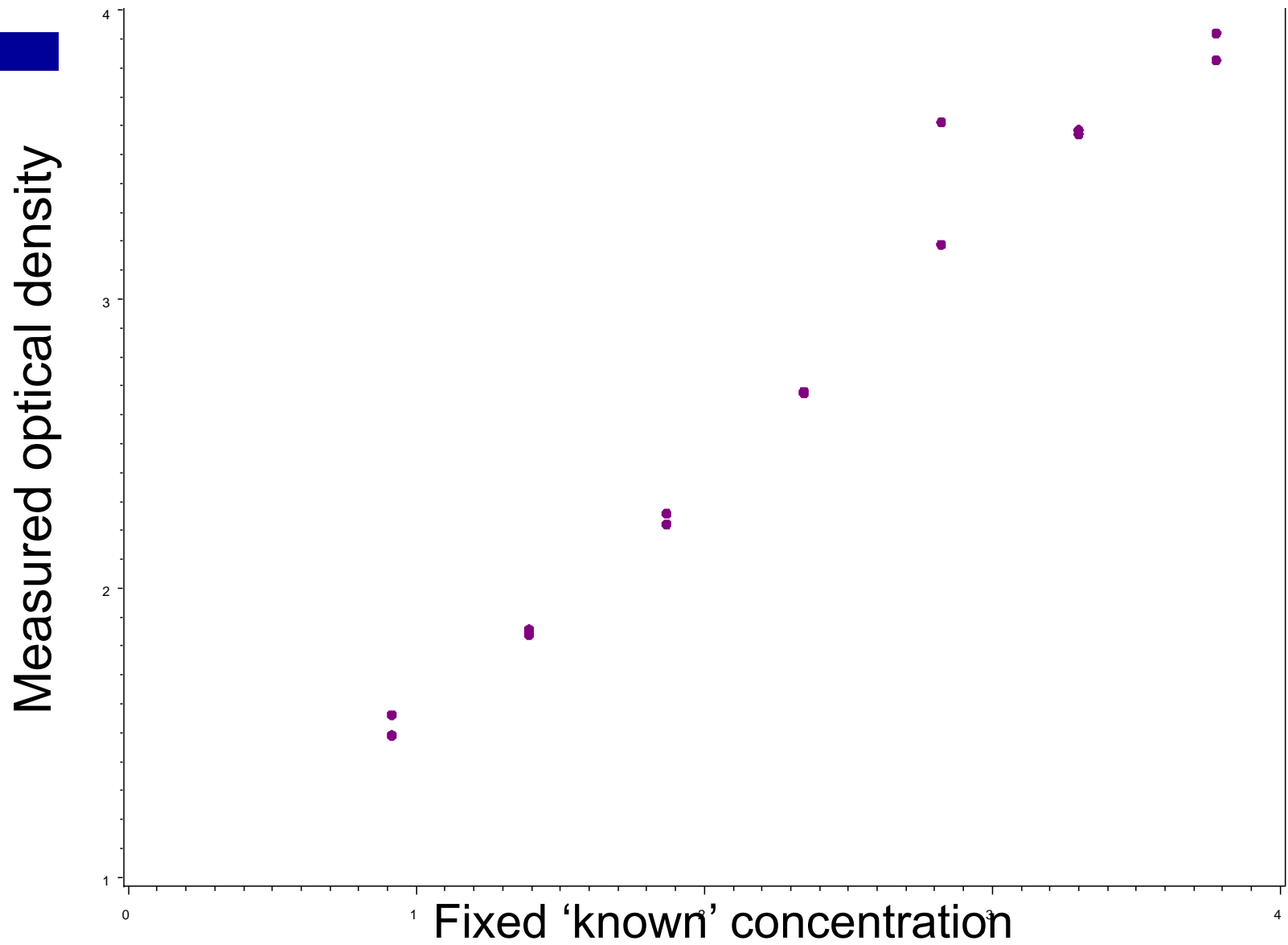
# Batch 6 calibration curve - GCSF



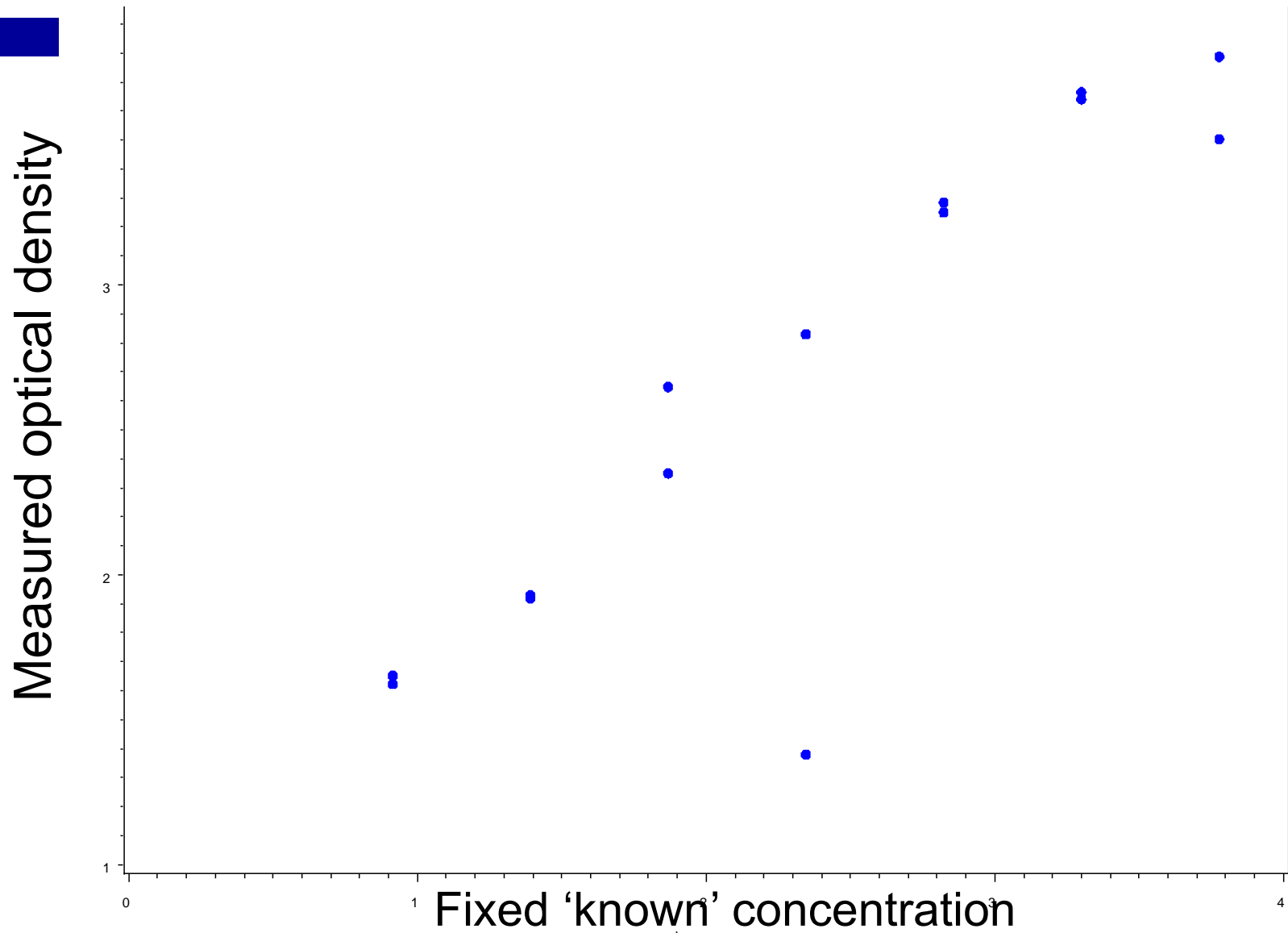
# Batch 9 calibration curve - GCSF



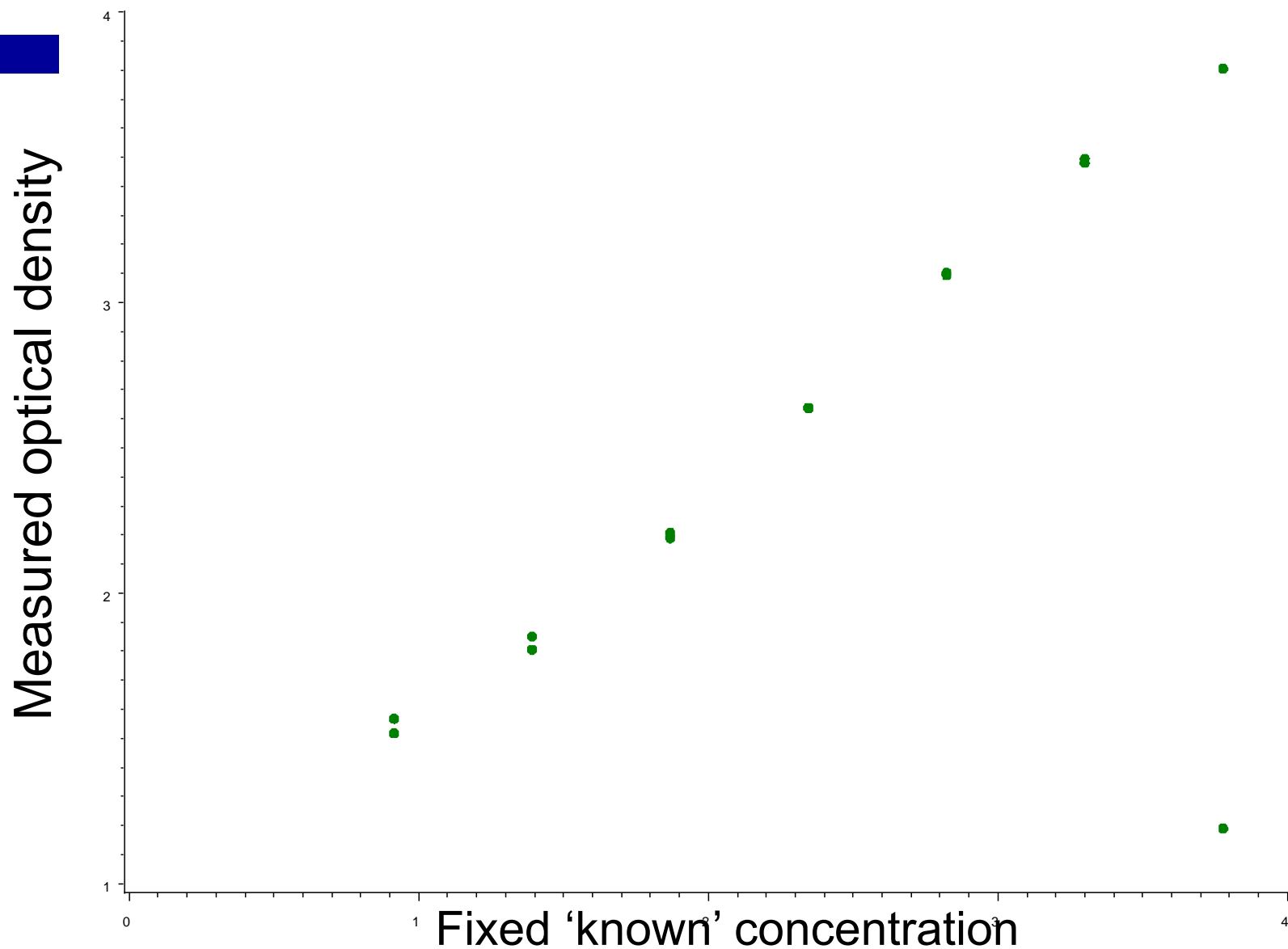
# Batch 10 calibration curve - GCSF



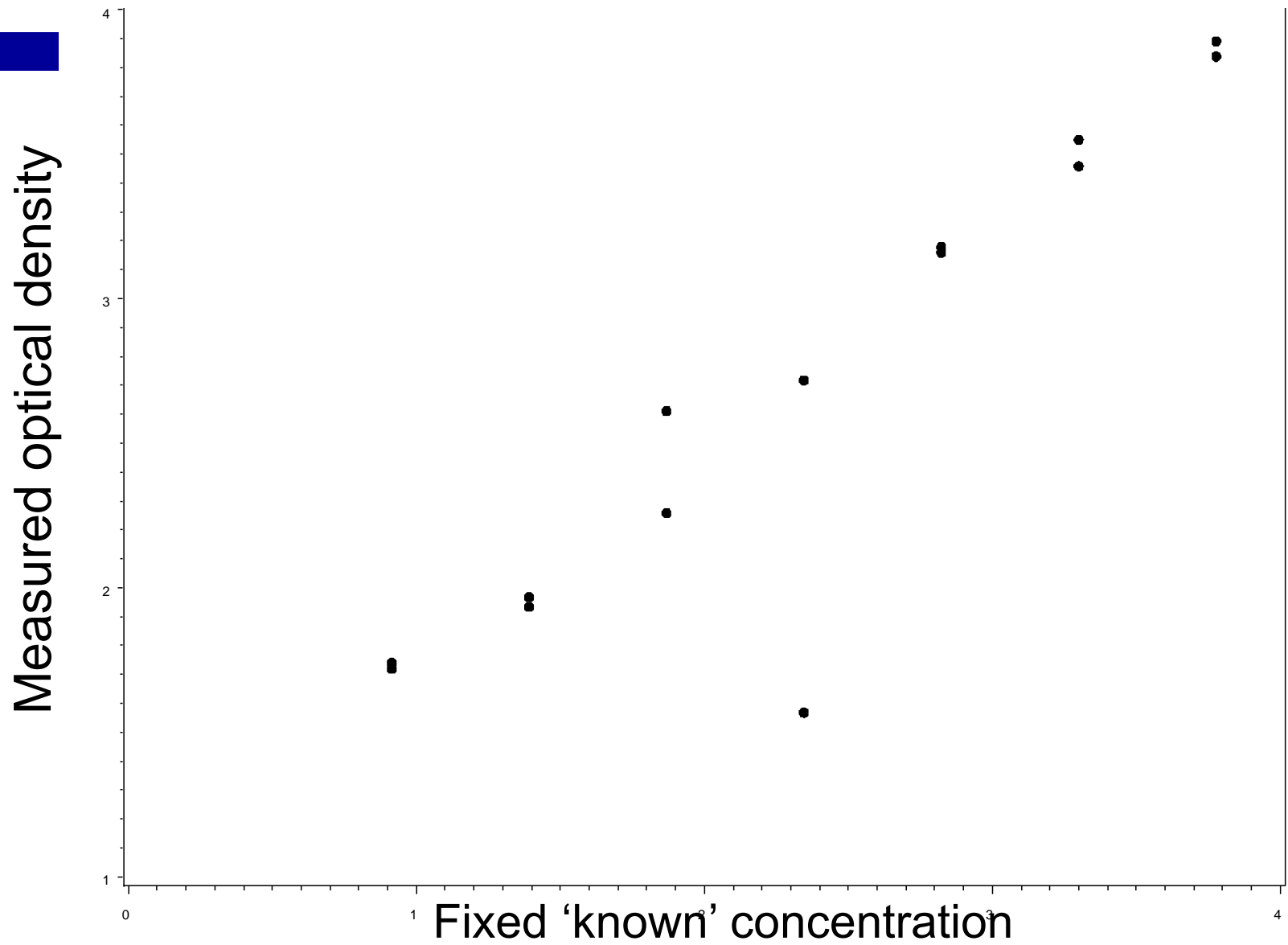
# Batch 21 calibration curve - GCSF



# Batch 22 calibration curve - GCSF

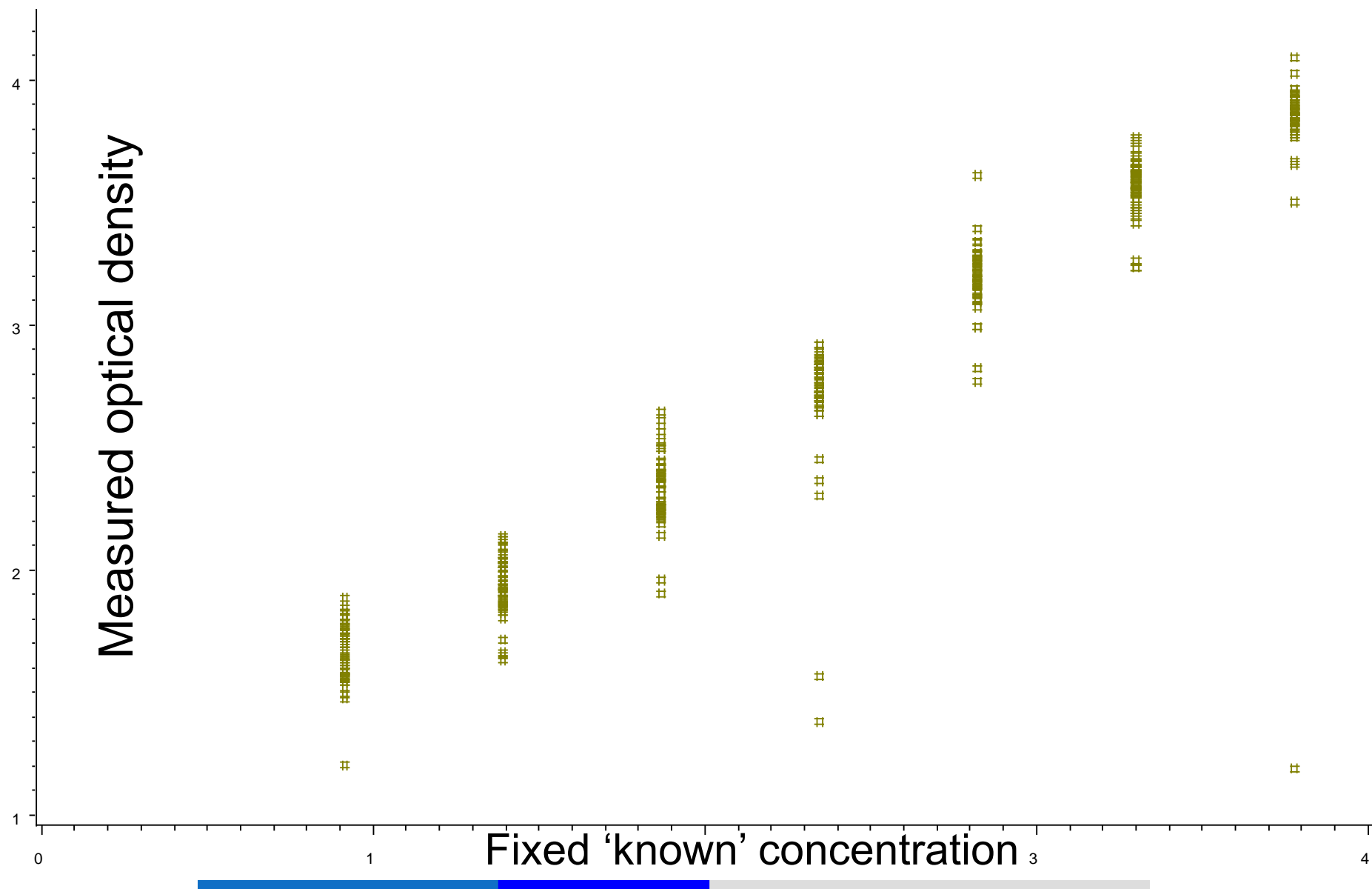


# Batch 24 calibration curve - GCSF





All calibration data (in log10)



# Effect of outliers on logistic regression results

Calibration models		As observed		Outliers removed	
		AOR	95%CI	AOR	95%CI
<i>Forward</i>					
Collapsed	Linear	0.34	(0.13, 0.90)	0.27	(0.10, 0.73)
Batch specific	Linear	0.73	(0.46, 1.17)	0.60	(0.33, 1.10)
Mixed model	Linear	0.67	(0.39, 1.14)	0.56	(0.30, 1.06)
Collapsed	Curvilinear	0.21	(0.05, 0.84)	0.25	(0.09, 0.71)
Batch specific	Curvilinear	0.81	(0.57, 1.16)	0.98	(0.93, 1.02)
Mixed model	Curvilinear	~	~	~	~
<i>Reverse</i>					
Collapsed	Linear	0.37	(0.15, 0.91)	0.28	(0.11, 0.73)
Batch specific	Linear	0.63	(0.37, 1.11)	0.58	(0.32, 1.07)
Mixed model	Linear	0.43	(0.19, 0.94)	0.53	(0.27, 1.02)
Collapsed	Curvilinear	0.37	(0.15, 0.91)	0.29	(0.11, 0.74)
Batch specific	Curvilinear	0.86	(0.53, 1.41)	0.67	(0.38, 1.16)
Mixed model	Curvilinear	0.50	(0.27, 0.95)	~	~

# Simulation study

1. Generate dataset with: true biomarker concentration, true effect on risk; overall relation between concentration and RFU; batch variability, and; occasional outliers
2. Simulate calibration experiments to estimate RFU – concentration relation according to each approach
3. Assess bias and variance of estimators from risk models

# Simulation study the biomarker

Biomarker:

$\exp(X \sim N(5,1))$

Miscarriage risk:

$OR = 1.05, 1.15 \text{ or } 1.65$

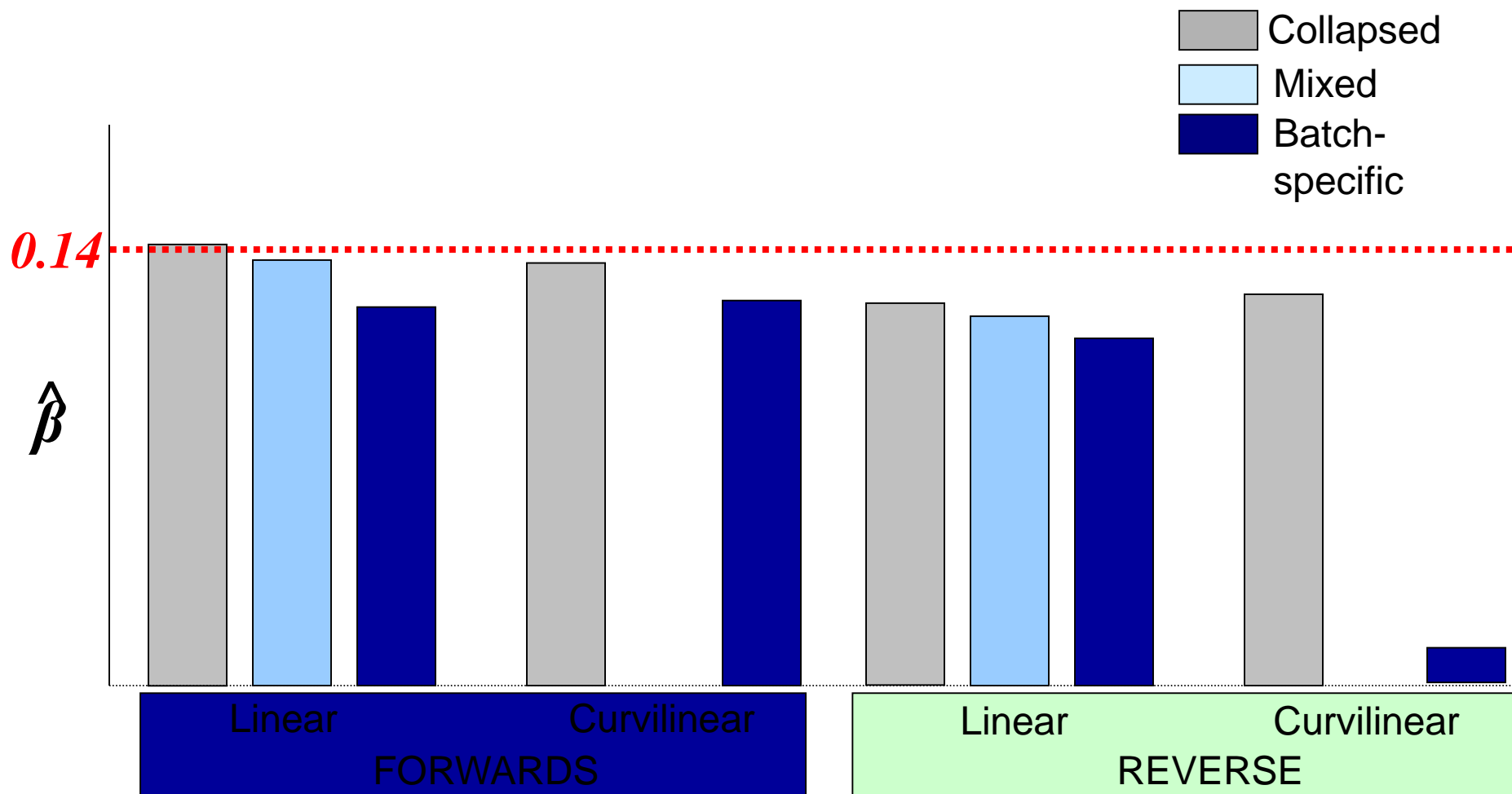
$\theta = \{0.05, 0.14, 0.50\}$

Conc. and OD:

OD determined through  
a single function

# Summary of simulation study results

Comparison of shape, model for  $\beta = 0.14$



# Conclusions

---

- Underestimation of effects due to calibration has implications for complex disease epidemiology
- Use of conventional batch-specific approaches performed poorly
  - Greatest bias to estimates in simulations
  - Most prone to loss of data for batches with failure of some calibration points