

CASSS

BIOASSAYS
SCIENTIFIC APPROACHES AND REGULATORY STRATEGIES 2018

16 – 17 April 2018

Minimization of Bioassay Test Result Shift When Implementing New Reference Standard Batches

Dr. Matthew Borer, Kim Danccheck, Dr. Paul Faya, Dr. Kristi Griffiths, David Lytle, Dr. Bhavin Parekh, Dr. Meng Zhao*

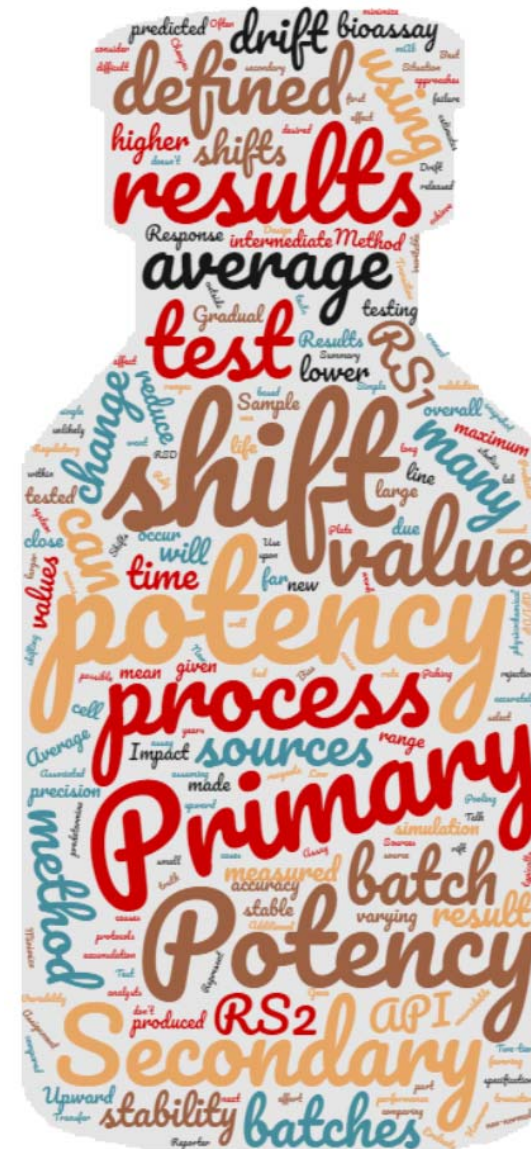
**Sr. Research Advisor*

*Corporate Reference Standards Organization
Eli Lilly and Company*

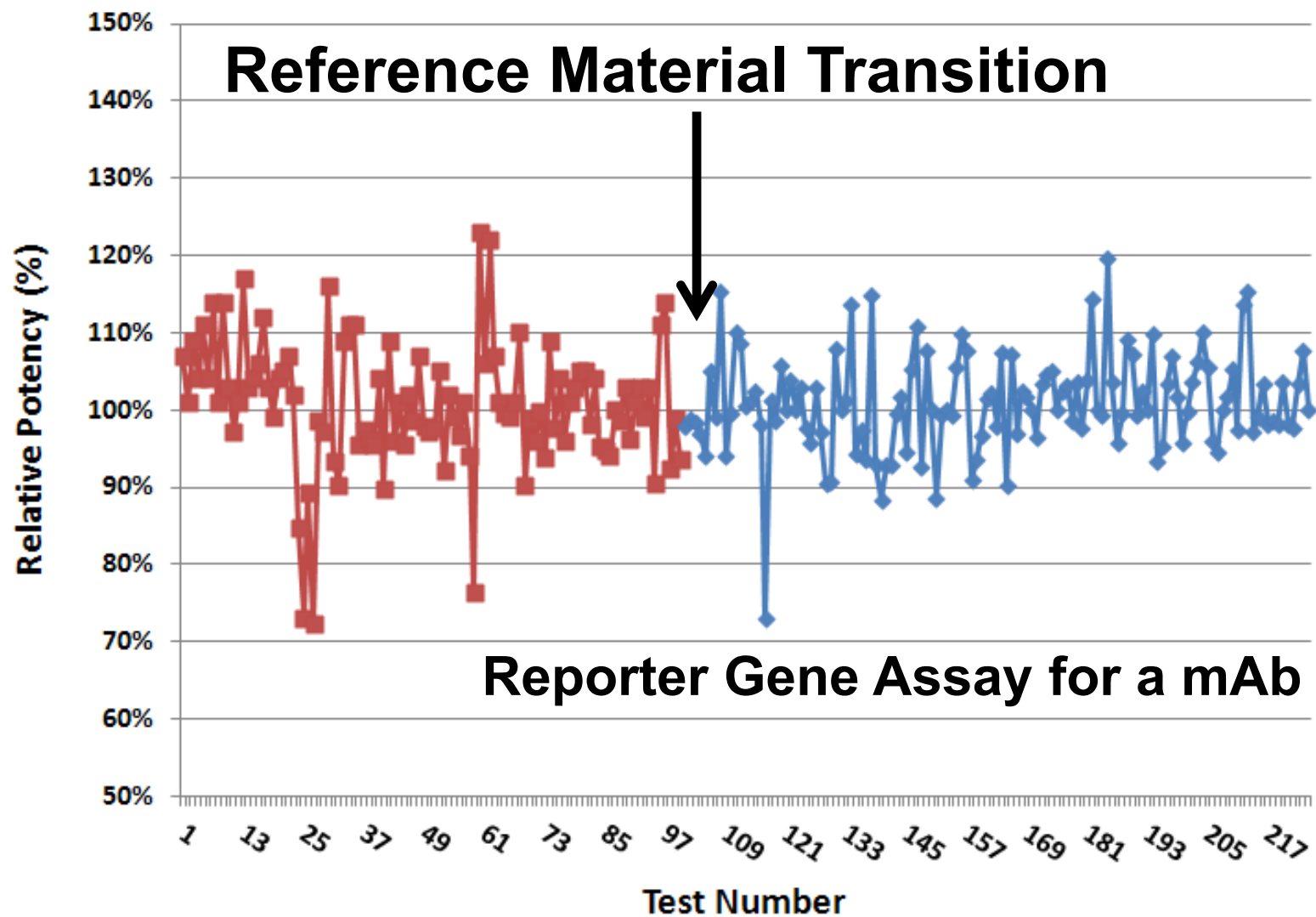
Lilly

Talk Outline

- ◆ Shift (and drift) defined
- ◆ Sources of shift
- ◆ Potency assignment approaches
- ◆ Calculation of shift
- ◆ Shift simulation
- ◆ Recommendations



Ideal Situation: No Drift, No Shift in test sample results over time



What is Shift?

- ◆ The overall average of test results changes upon implementation of a new reference material
- ◆ Caused by
 - Differences in the relative value assignment of each RS batch
 - Variability
 - Bias

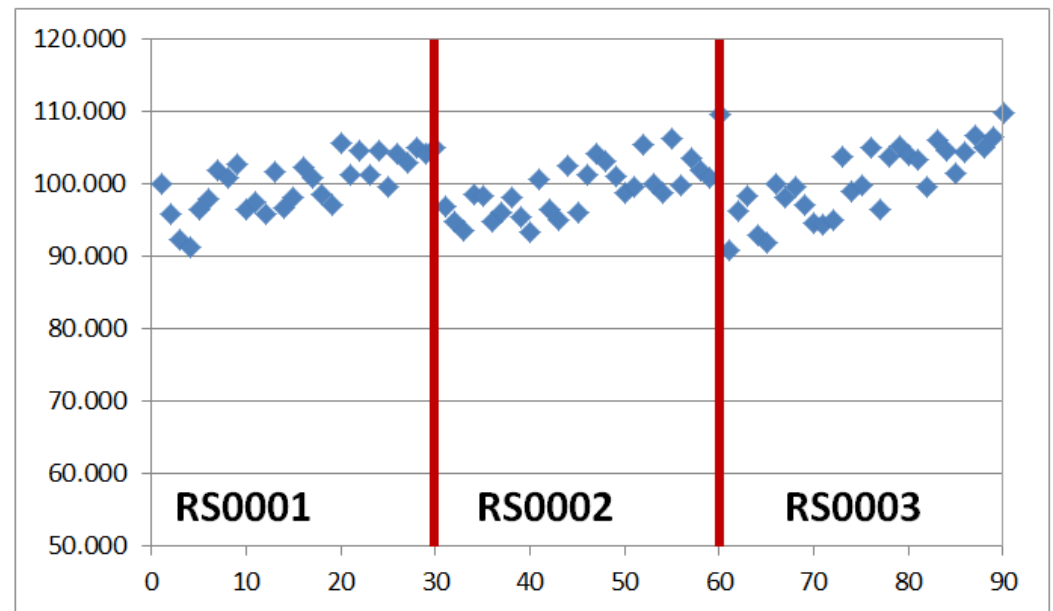
In this hypothetical example, each RS is a Secondary RS with variability in the defined potency, varying around a “true value” of 100%



What is Drift?

- ◆ Test results that gradually change over the life of a single batch of reference material
- ◆ Caused by
 - Reference material degradation (drift upward)
 - Gradual method changes
 - Gradual process changes

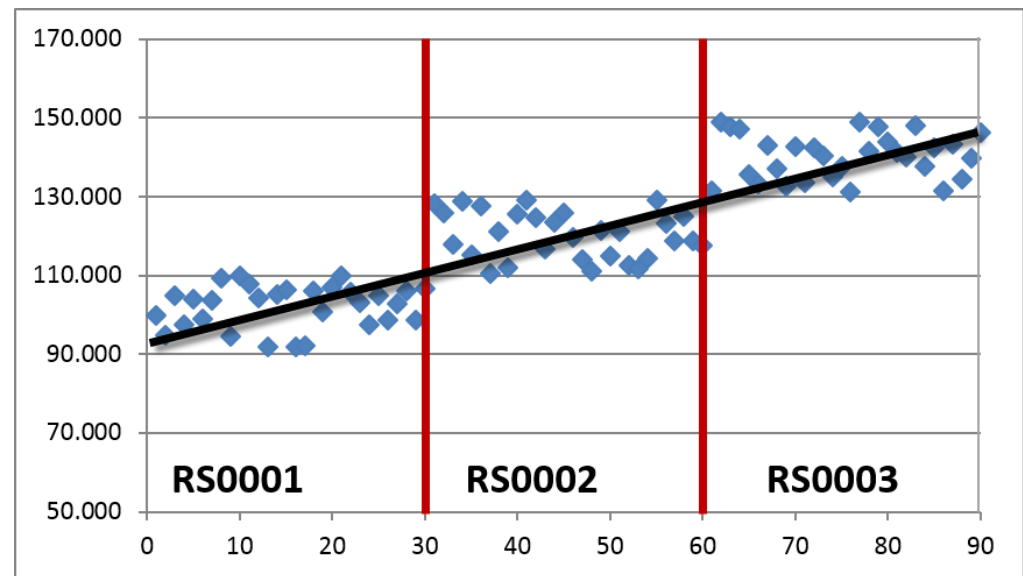
In this hypothetical example, each RS is a Secondary RS, each of which degrades 15% over its life. The Secondary RS batches are accurately defined against a stable Primary RS.



What else causes Drift?

- ◆ An accumulation of shifts from one reference standard batch to the next
- ◆ Caused by
 - Coincidentally shifting up at each assignment due to method variability
 - Degrading Primary RS

In this hypothetical example, each RS is a stable Secondary RS that is defined against a Primary RS that is degrading 20% between each Secondary RS (with no change being made to the Primary RS defined potency)

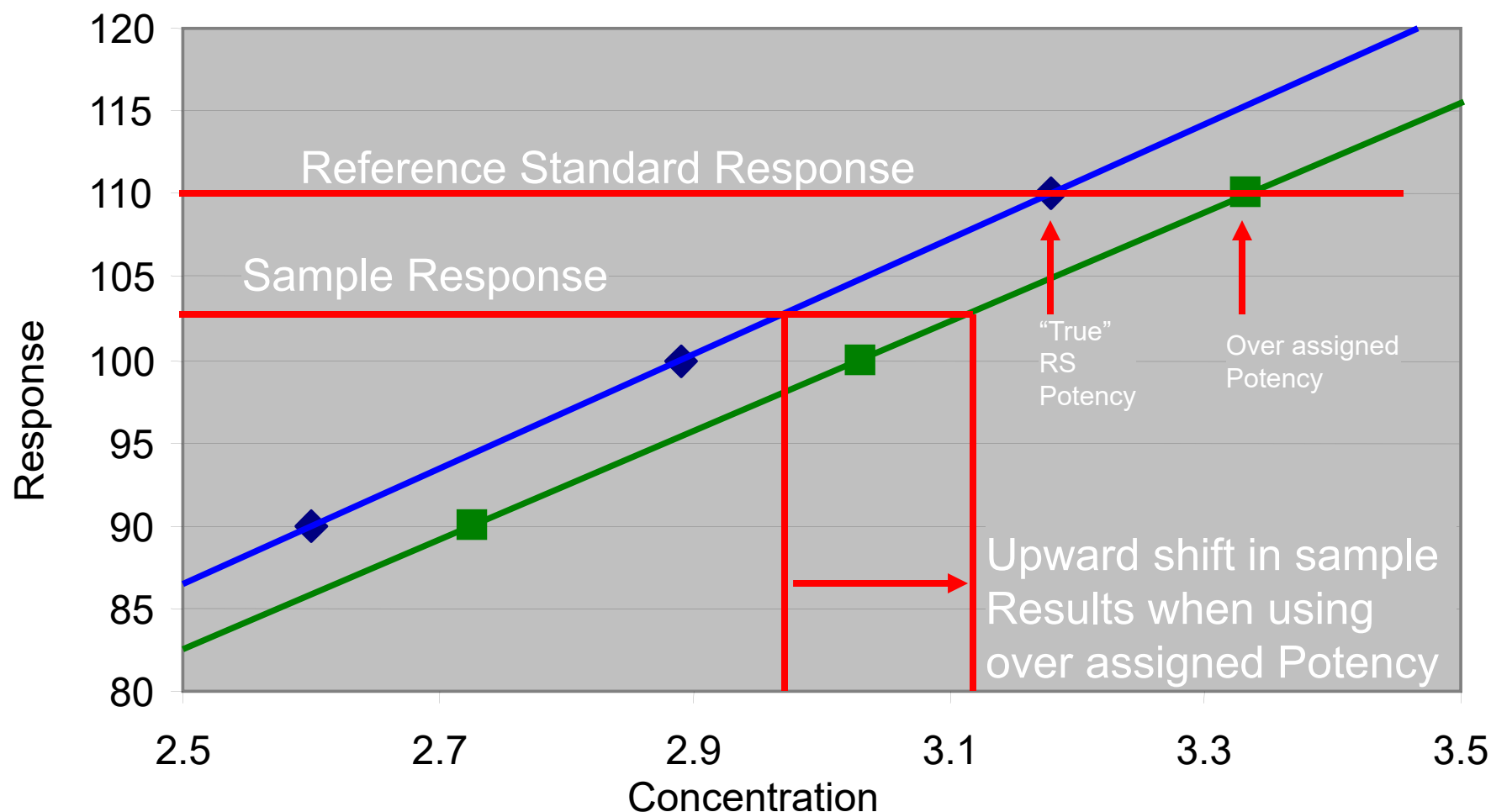


Definitions

- ◆ True Potency - The potency of a material if it could be measured with perfect accuracy and no variability.
- ◆ Assigned Potency – The potency of a material that was measured and calculated.

Assigned Potency is different from True Potency due to lack of perfect accuracy and method variability.

Assignment $\neq$ True Potency: Impact on Sample Results



Over/Under Assignment

- ◆ Over-assigned = When the assigned value is higher than the “true” value

An over-assigned reference standard yields higher test results than an perfectly assigned reference standard

- ◆ Under-assigned = When the assigned value is lower than the “true” value

An under-assigned reference standard yields lower test results than an perfectly assigned reference standard

What is Shift?

- ◆ The overall average of test results changes upon implementation of a new reference material
- ◆ Caused by
 - Differences in the relative value assignment of each RS batch (***RS batches not all assigned at true potency***)
 - Variability
 - Bias

In this hypothetical example, each RS is a Secondary RS with variability in the defined potency, varying around a “true value” of 100%



Shift Magnitude – Example Calculation

◆ Reference Standard 1

- RS1 was produced using API material with 102% True Potency
- RS1 was tested as a sample using the Primary RS. The average result was 99% so RS1 was given an Assigned Potency of 99%.
- Therefore, RS1 is 3% under-assigned and yields test results that are 3% lower than a “perfect” RS.

◆ Reference Standard 2

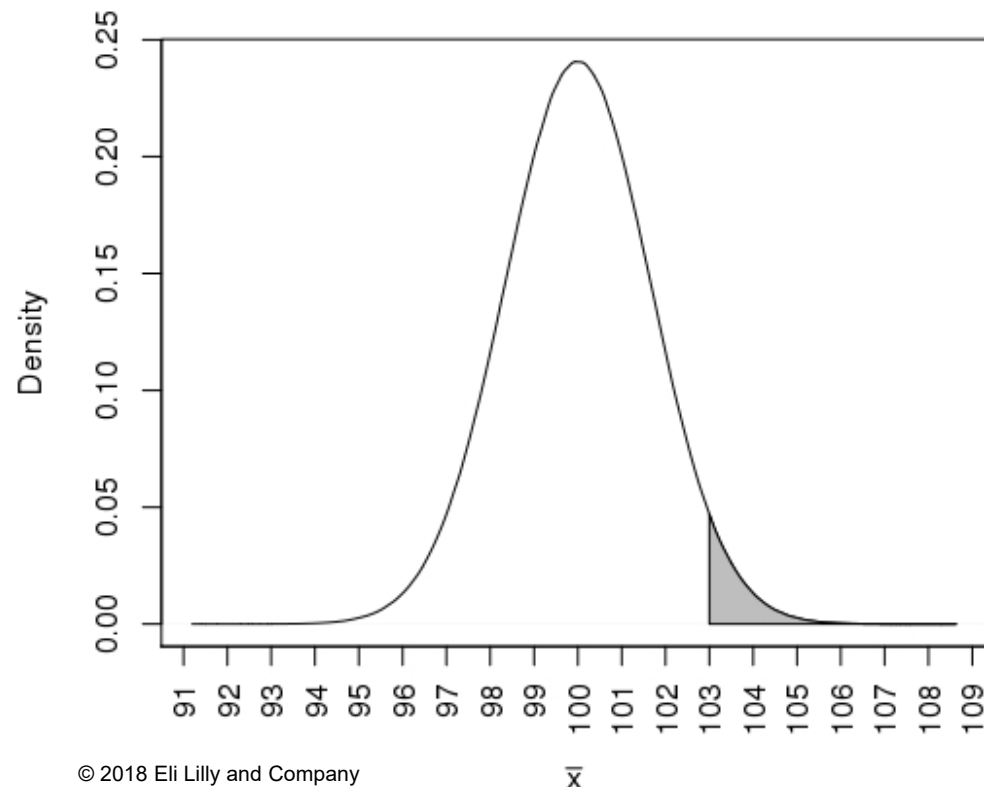
- RS2 is produced using API material with 98% True Potency
- RS2 is tested as a sample using the Primary RS. The average result is 101% so RS2 is given an Assigned Potency of 101%.
- Therefore, RS2 is 3% over-assigned and yields test results that are 3% higher than a “perfect” RS.

The transition from RS1 to RS2 will result in a 6% upward shift!

Shift Probability

- ◆ By averaging many independent measurements, we can reduce the likelihood that the measured value is far from the true value
- ◆ For example, the probability that the reference standard will be 3% or more over-assigned is about 3.5% (assuming 100% true potency, bioassay variability of 5.5%, and averaging 11 independent measurements)

$$\begin{aligned}\bar{x} &\sim N\left(\mu_M, \frac{\sigma}{\sqrt{n}}\right) \\ &= N\left(100, \frac{5.5}{\sqrt{11}}\right) \\ \Pr(\bar{x} \geq 103) &= 0.035\end{aligned}$$



Shift Probability as a function of replication (5.5% IP method)

Number of replicates	Probability of a 6% shift (%)
1	8.57
3	2.97
6	0.82
9	0.26
12	0.09
15	0.03



Generating 11 independent measurements typically requires multiple weeks and large expenditures.

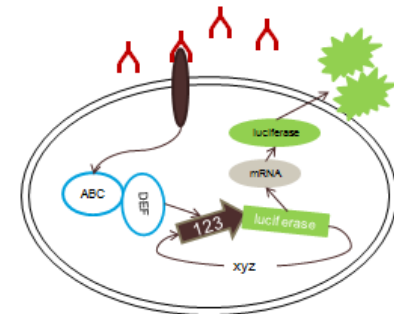
Testing is infrequent and can't easily be build in to lab capacity plans.

Avoiding shift is expensive and complicated!



Non-normal Sources of Shift

- ◆ Averaging many results can reduce the probability of a shift
- ◆ However, estimates of bioassay intermediate precision are typically a snapshot in time compared to infrequent RS batch replacements
 - Method validation studies don't incorporate all sources of variability that occur over many years
 - Method drift
 - Changes to critical reagents
 - Inhomogeneity of the cell line
 - Epigenetic changes to the cell line
 - Transfer to new lab
 - New analysts
 - Different equipment



Averaging more results doesn't eliminate all sources of shift!

Summary so far...

- ◆ Shift and drift are bad...we want consistent measurement of potency over time
- ◆ There are many sources of shift and drift, some of which are unavoidable
- ◆ Simple variability in the measurement causes some shift, although large shifts can be made unlikely by averaging
- ◆ There are many non-normal sources that can introduce RS shifts which are difficult to account for because of the long time between RS batch replacements

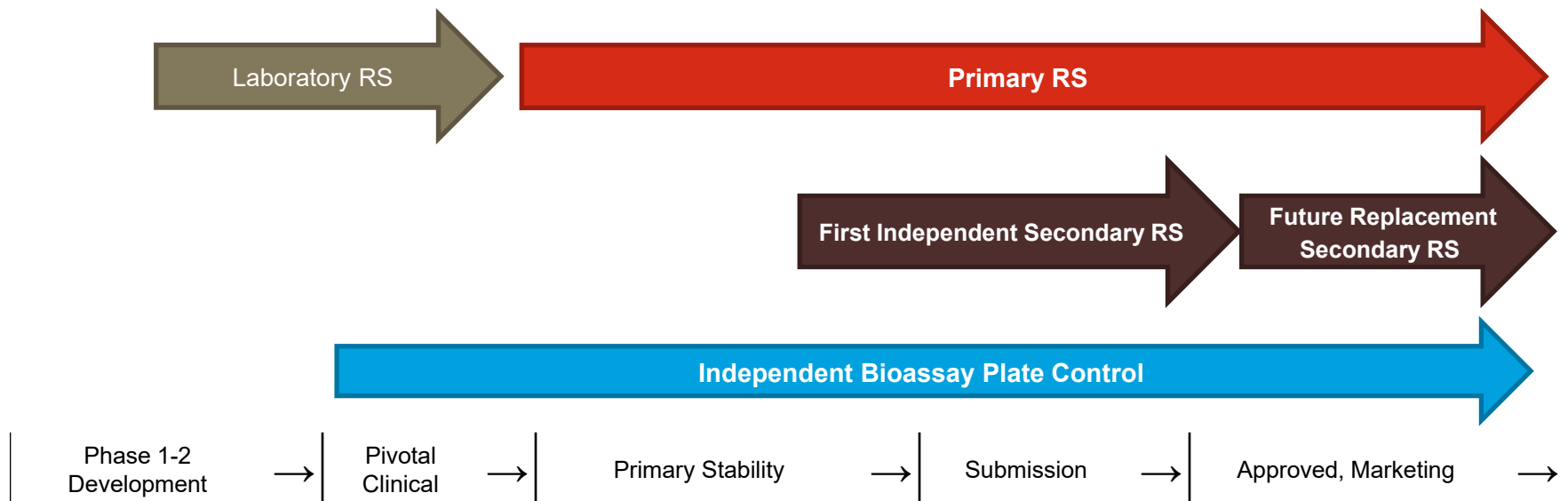


Minimize Shift and Drift

- ◆ Two-tier system
 - Primary and Secondary RS
- ◆ Effective means to **monitor stability** of RS batches
 - for changes that affect potency
- ◆ **Design characterization protocols** to minimize shift between RS batches

Monitoring Stability

- ◆ Impossible to correlate potency to physicochemical tests
- ◆ Rely on “parallel stability”
 - comparing values that should all independently be constant
 - Primary RS, Secondary RS, Plate Control, manufacturing process, API/DP stability, bioassay performance



Characterization Protocol Design

- ◆ Primary RS (and first Development RS) assigned 100% - defines “truth” through a highly representative material
- ◆ When defining potency by comparative testing
 - Best effort to reduce the variability of the potency assay
 - Perform sufficient replication to achieve desired uncertainty in the average value
 - Represent all sources of variability between independent measurements

Characterization Protocol Design: Secondary RS Batches

- ◆ Assign the Average
 - Assign the Secondary RS potency using the average value of test results against the Primary RS
- ◆ Assign 100%
 - Assign the Secondary RS potency at 100% if the average value of test results against the Primary RS is within a defined range

In both cases, if the average test result against the Primary RS is outside of a defined range (95 – 105% in the following work), consider rejection of the Secondary RS because it is not representative of the Primary RS and the manufacturing process.

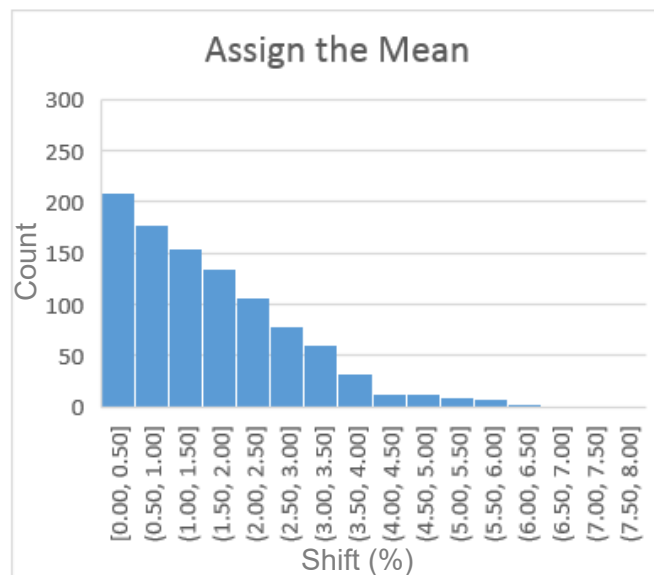
How to select the approach

- ◆ Assign at 100%
 - When the manufacturing process variability is small relative to the method variability
- ◆ Assign the Average
 - When the manufacturing process variability is large relative to the method variability

Evaluate by simulation

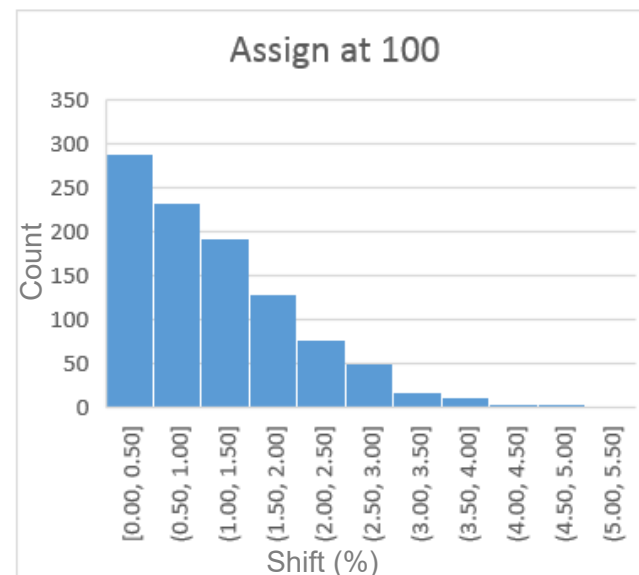
Low Process Variability

API Mean True Potency 100
 API True Potency RSD% 1
 Bioassay Variability RSD% 5
 Replication 10



Failure Rate: 0.6%

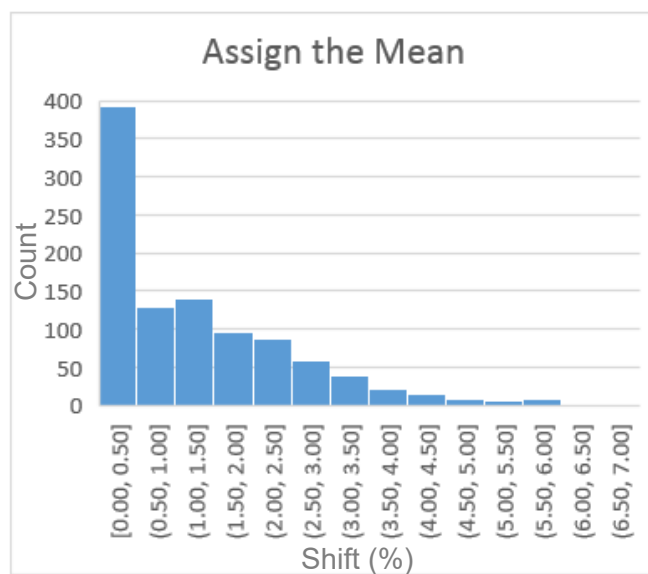
Max Shift 7.9
 Average Shift 1.6
 Median Shift 1.4



Max Shift 5.0
 Average Shift 1.1
 Median Shift 1.0

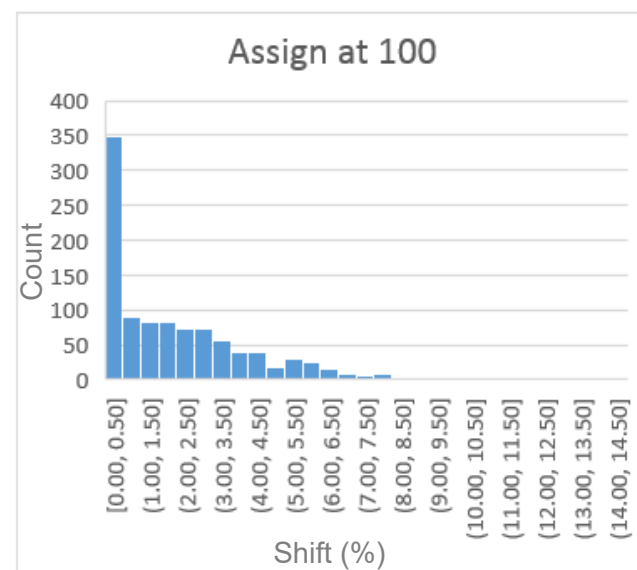
High Process Variability

API Mean True Potency 100
 API True Potency RSD% 3
 Bioassay Variability RSD% 5
 Replication 10



Failure Rate: 13.3%

Max Shift 6.9
 Average Shift 1.6
 Median Shift 1.4



Max Shift 14.3
 Average Shift 2.5
 Median Shift 2.1

Optimal Design

Assign 100	Median Shift
$\Delta 0.5 - 2.8\%$	Maximum Shift
(0.0)	Measured value >105, <95 (%)

- ◆ Technique with the smallest shift, varying process and method RSD values

Process %RSD \ Method %RSD	0.5	1	1.5	2	2.5	3
3 (n=6)	Assign 100 $\Delta 0.5 - 2.8\%$ (0.0)	Assign 100 $\Delta 0.9 - 4.8\%$ (0.0)	Average $\Delta 1.2 - 5.2\%$ (0.6)	Average $\Delta 1.2 - 6.1\%$ (3.8)	Average $\Delta 1.2 - 5.3\%$ (6.3)	Average $\Delta 1.1 - 5.6\%$ (13.2)
4 (n=8)	Assign 100 $\Delta 0.5 - 2.8\%$ (0.1)	Assign 100 $\Delta 1.0 - 5.4\%$ (0.1)	Average $\Delta 1.2 - 6.5\%$ (1.2)	Average $\Delta 1.4 - 5.8\%$ (4.8)	Average $\Delta 1.3 - 6.7\%$ (7.9)	Average $\Delta 1.2 - 6.3\%$ (14.6)
5 (n=10)	Assign 100 $\Delta 0.5 - 2.5\%$ (0.4)	Assign 100 $\Delta 1.0 - 5.2\%$ (0.5)	Assign 100 $\Delta 1.4 - 6.7\%$ (2.4)	Average $\Delta 1.5 - 7.0\%$ (4.3)	Average $\Delta 1.3 - 8.0\%$ (8.3)	Average $\Delta 1.5 - 8.0\%$ (14.3)
6 (n=13)	Assign 100 $\Delta 0.5 - 2.4\%$ (0.3%)	Assign 100 $\Delta 0.9 - 5.4\%$ (0.9)	Assign 100 $\Delta 1.6 - 6.1\%$ (2.9)	Average $\Delta 1.5 - 7.8\%$ (4.4)	Average $\Delta 1.4 - 7.0\%$ (10.1)	Average $\Delta 1.4 - 7.8\%$ (15.5)
7 (n=16)	Assign 100 $\Delta 0.5 - 2.2\%$ (0.4)	Assign 100 $\Delta 1.0 - 4.4\%$ (0.6)	Assign 100 $\Delta 1.4 - 6.8\%$ (3.3)	Average $\Delta 1.6 - 7.7\%$ (6.8)	Average $\Delta 1.8 - 7.7\%$ (10.2)	Average $\Delta 1.6 - 7.2\%$ (15.2)
8 (n=19)	Assign 100 $\Delta 0.5 - 2.4\%$ (0.8)	Assign 100 $\Delta 0.9 - 5.9\%$ (2.2)	Assign 100 $\Delta 1.5 - 7.4\%$ (3.8)	Average $\Delta 1.6 - 6.9\%$ (6.0)	Average $\Delta 1.7 - 7.7\%$ (14.4)	Average $\Delta 1.5 - 7.6\%$ (15.1)

Additional Considerations

- ◆ Picking a source batch based on sufficient testing can predetermine that the RS batch has a potency close to the process mean
- ◆ Pooling many API batches can increase confidence that the RS candidate is close to the process mean
- ◆ Both have the same effect as lowering the process variability, favoring the Assign at 100% approach

Impact of inevitable shifts

- ◆ Ideally, the strategy is well defined, followed correctly, and a maximum shift can be predicted
- ◆ Predicted maximum shift should be included as part of the overall specification uncertainty budget
 - Often not considered because shift is so infrequent
- ◆ Shifts larger than predicted will occur!
 - Determine cause if possible
 - Investigate impact
 - Regulatory commitments
 - In-house acceptance ranges
 - Manufacturing strategy
 - Use of API released before the change to formulate DP after the change

Thank You!

