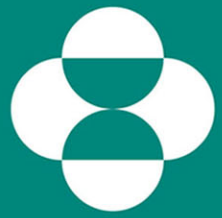
Validation of Rapid Microbiological Methods:

Statistical and Regulatory Challenges

BMS-ANed & PSDM Fall meeting
23 Nov 2018, Wageningen



MSD
INVENTING FOR LIFE

Pieta IJzerman-Boon
Center for Mathematical Sciences, MSD

Content

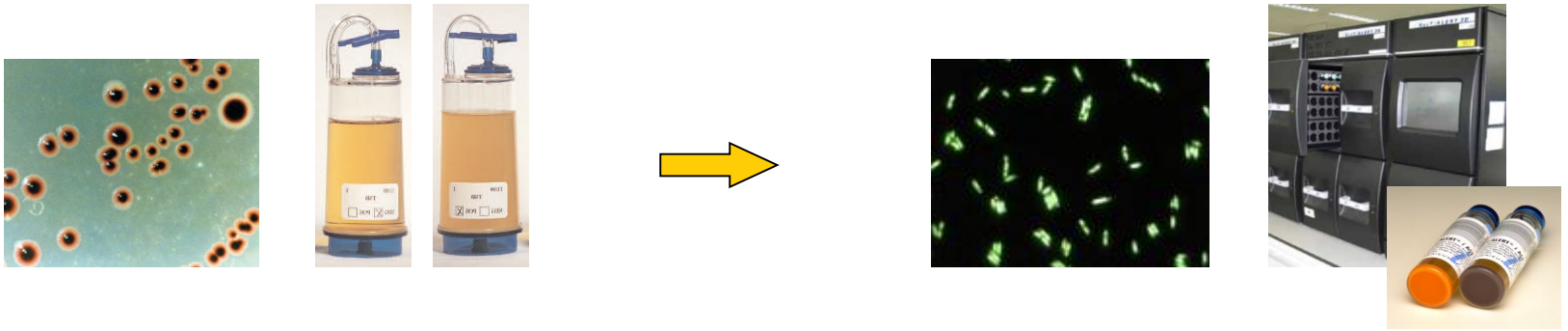
- Introduction
- Non-inferiority for qualitative tests according to USP <1223>
- Statistical model for detection of microorganisms
- Simulations
- Conclusions

Introduction
Different test methods

- Analytical test methods are being developed during development of a new product
  - Identification tests
  - Impurities (quantitative and limit tests)
  - Assays (quantitative tests for the active ingredient or other components of drug substance or drug product)
- Tests are used for
  - Release and stability testing of the product
  - In-process tests on intermediates
- Microbiological methods are used to test for the presence or occurrence of microorganisms in product, process, or environment

Introduction
Rapid Microbiological Methods (RMM)

- Alternative rapid microbiological methods (RMMs) are developed to replace conventional growth-based methods



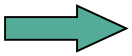
  - Reduce time to result from ≥ 2 weeks to 0-5 days
  - In-process controls, Root cause investigations, Release
  - Improve lab efficiency
  - Cost savings (e.g. when batch can be saved)
    - Early detection of contaminations and reliable counting methods may save $millions/year

Introduction
Validation

- The objective of validation is to demonstrate that the method is suitable for its intended purpose (ICH Q2(R1))
  - Do we need a limit of detection of one?
  - What if a few organisms are not always counted?
  - Can we accept false positives? And how many?
  - How precisely should we determine the capability?
  - Should RMM be better than compendial?
- Microbiological guidelines EP 5.1.6 & USP <1223>
  - Are not fully developed and aligned
  - Accuracy, precision, limit of detection (LOD), specificity
  - Suggest that RMM should be equivalent or non-inferior to conventional method

Introduction
Validation Issues

- Validation RMMs more complex than analytical methods
  - Living organisms, which may be sensitive to conditions
  - Different species, which will respond differently
  - Impossible to spike precisely – lack of ref standards
  - False positives may mask false negatives for non-growth-based methods



- Performance cannot be observed directly from the results
- Statistics needed!

Introduction
Validation Issues

- Ideal Microbiological Experiments:

Qualitative Tests

– Repeated blank samples for specificity

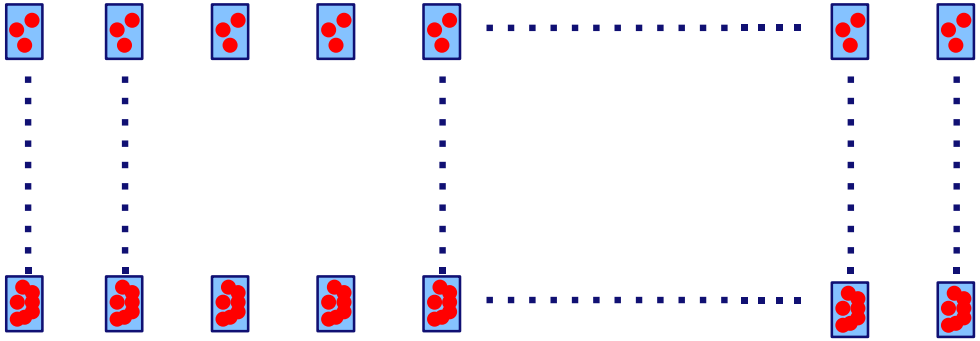
→ False positive rate

– Repeated samples with one microorganism for detection limit

→ False negative rate

Quantitative Tests

– Repeated samples with higher numbers of microorganisms for accuracy



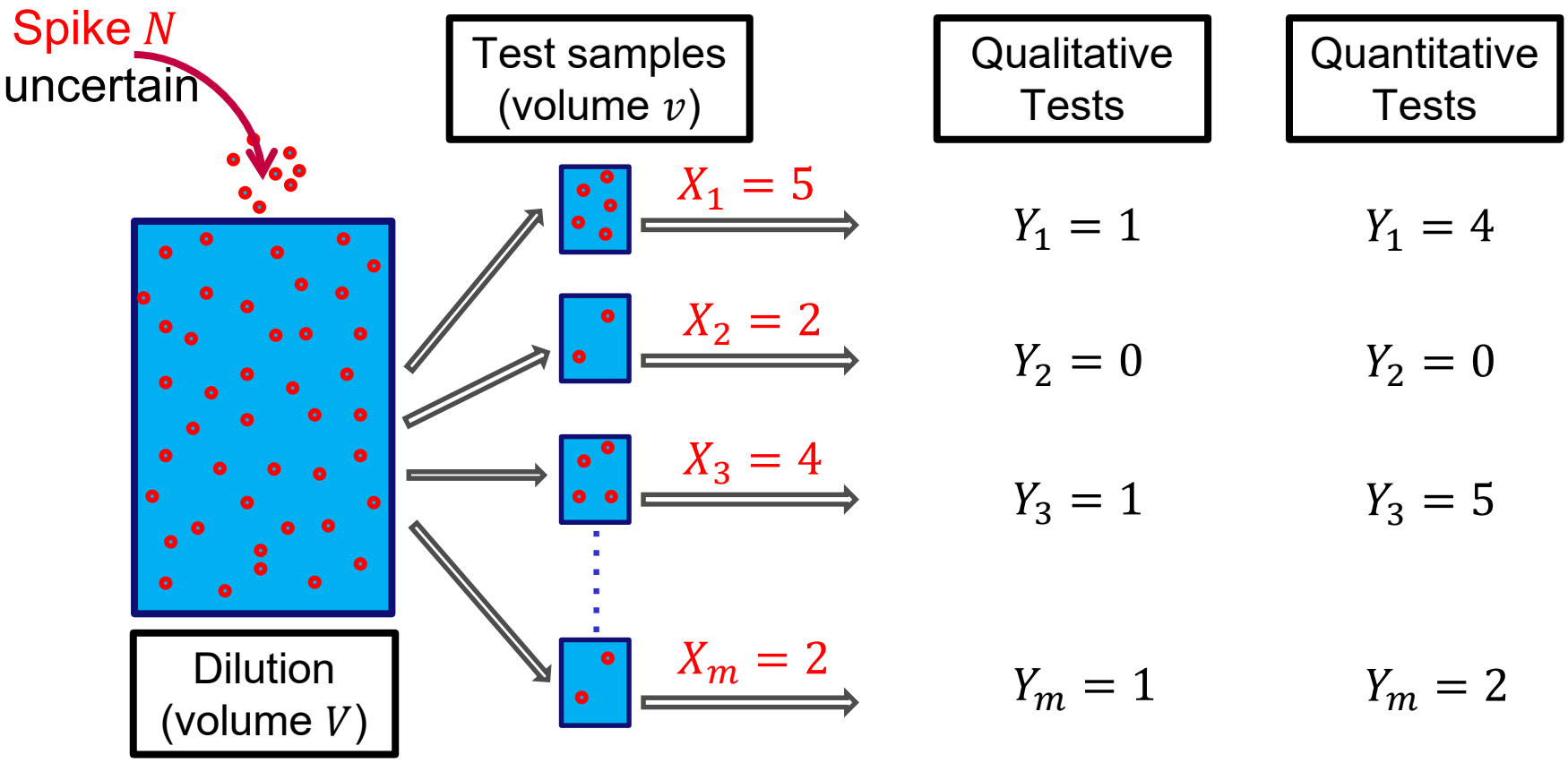
→ Bias – Spike 1

→ Bias – Spike m

Traditional statistical analysis methods can be applied to the ideal microbiological experiment

Introduction
Validation Issues

- Real Microbiological Experiments:



True number of organisms X in test sample is Binomial$(N, v/V)$ or approximately Poisson(vN/V)

Non-inferiority according to USP <1223>
Approach 1

Approach 1: Non-inferiority on positive rates

- Use spike level for which 50-75% of samples is tested positively with compendial method
- Hypotheses: H_0: $p_A/p_C \leq r_0$ versus H_1: $p_A/p_C > r_0$
  - Non-inferiority margin: $r_0 = 0.8$
- Reject H_0 if (Farrington & Manning, 1990):

$$\frac{\hat{p}_A - r_0\hat{p}_C}{\hat{w}_0^{1/2}} > z_{1-\alpha}$$

  - with $\hat{p}_A$ and $\hat{p}_C$ the proportions of positive samples for alternate and compendial method
  - with $\hat{w}_0$ the MLE of $\mathrm{var}(\hat{p}_A - r_0\hat{p}_C)$ under H_0

Non-inferiority according to USP <1223>
Approach 2

Approach 2: Non-inferiority on most probable numbers (MPN)

- The MPN estimates or quantifies the number of organisms in a suspension based on qualitative (pos/neg) results of samples taken from (dilutions of) the suspension (Cochran, 1950)

- Hypotheses: $H_0: \mu_A - \mu_C \leq \log(r_0)$ vs. $H_1: \mu_A - \mu_C > \log(r_0)$
  - with μ_A and μ_C the mean MPNs in log scale
- Use t-test for non-inferiority: reject H_0 if the one-sided 95% LCL for $\mu_A - \mu_C$ exceeds $\log(r_0)$

Non-inferiority according to USP <1223>
MPN using one dilution

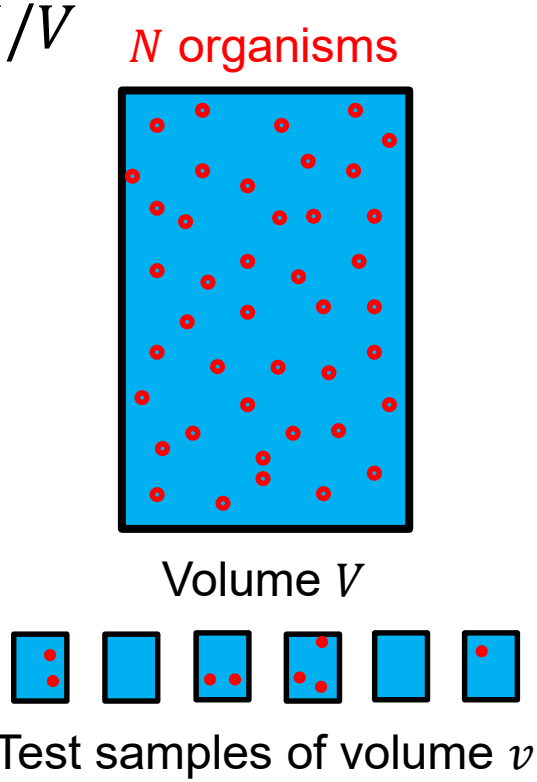
- One suspension with N organisms in volume V mL
- For a test sample with volume v, the mean bacterial density per test sample is $\lambda = vN/V$
- The probability that it is contaminated is

$$p = 1 - \exp(-\lambda)$$

  assuming the number of organisms in a sample follows Poisson
- Hence, the MPN becomes

$$\hat{\lambda} = -\log(1 - \hat{p}), \quad \hat{N} = (V/v)\,\hat{\lambda}$$

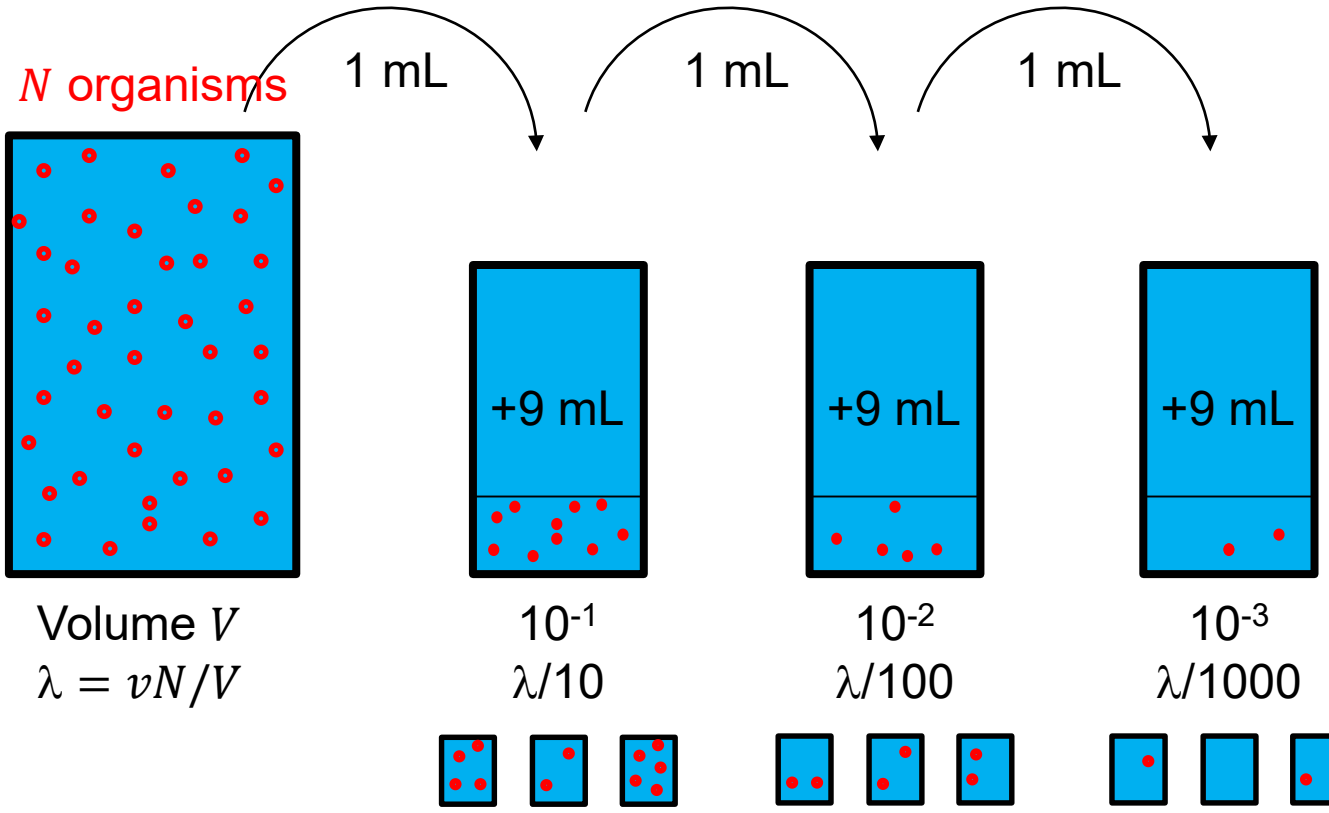
- Does not exist if all samples positive



Example: $\hat{\lambda} = -\log(1 - 4/6) = 1.1$

Non-inferiority according to USP <1223>
MPN using multiple dilutions

- Multiple dilutions to ensure both pos and neg samples
- Usually 3 or 5 dilutions (10- or 2-fold) with 3 or 5 replicates
- For multiple dilutions, no closed-form expression exists



Example:

3,3,2 positive

$\widehat{N} = 1100$

Non-inferiority according to USP <1223>
Implicit assumptions

Approach 2:

- Cochran, 1950:
  - The organisms are distributed randomly throughout the liquid. Thus the liquid is thoroughly mixed.
  - Each sample from the liquid, when incubated in the culture medium, is certain to exhibit growth whenever the sample contains one or more organisms.
- Gartright & Blodgett, 1996:
  - Random, unaggregated distribution of bacteria so that the number in a small unit follows Poisson
  - Each tube is independent of the others
  - Growth will ensue in a sterile tube with the introduction of one or more bacteria

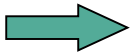
→

- Implicitly assumes LOD=1 – What if this is not the case?

Non-inferiority according to USP <1223>
Implicit assumptions

Approach 1:

- Non-inferiority claim would only hold for tested spike level
  - What if spike was too high?
  - What about other spikes?
- Binomial probabilities for all samples assumed to be same
  - Due to spiking and sampling variability, samples have different detection probabilities



- Number of positive test samples depends not only on microbiological test method, but also on the numbers of organisms in samples
  - Not controlled due to spiking and sampling variability

Statistical detection model

- Need a model that separates
  - Detection by the microbiological method
  - Spike variability and sampling process
- Classification of a qualitative test result

Test Result	True Number of Organisms $X = 0$	True Number of Organisms $X > 0$
$Y = 1$ $Y = 0$	Specificity	False Negative
	False Positive	Sensitivity: LOD

Note: USP/EP define specificity as the ability to detect a range of organisms (=sensitivity)

- So we need to look at the conditional detection probabilities

$$\pi(x) = P(Y = 1 | X = x)$$

Statistical detection model

- Binomial Mechanism (BM)

$$\pi(x) = 1 - (1 - \theta)^x$$

  - Each organism has a probability θ to be detected
  - This detection proportion is related to sensitivity/LOD
  - Seems reasonable for growth-based methods

- Can be extended with a false positive rate η for specificity (IJzerman-Boon & Van den Heuvel, 2015)

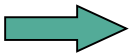
- However, we do not know x, so we cannot estimate $\pi(x)$, only the average over different samples

Statistical detection model

- If true number of organisms X is Poisson(λ), then the marginal probability of a positive sample is

$$p = 1 - \exp(-\theta\lambda)$$

- The positive rates (USP1) estimate this marginal probability
- The MPN estimator (USP2) only estimates λ if $\theta = 1$
- In general, only the product $\xi = \theta\lambda$ can be estimated



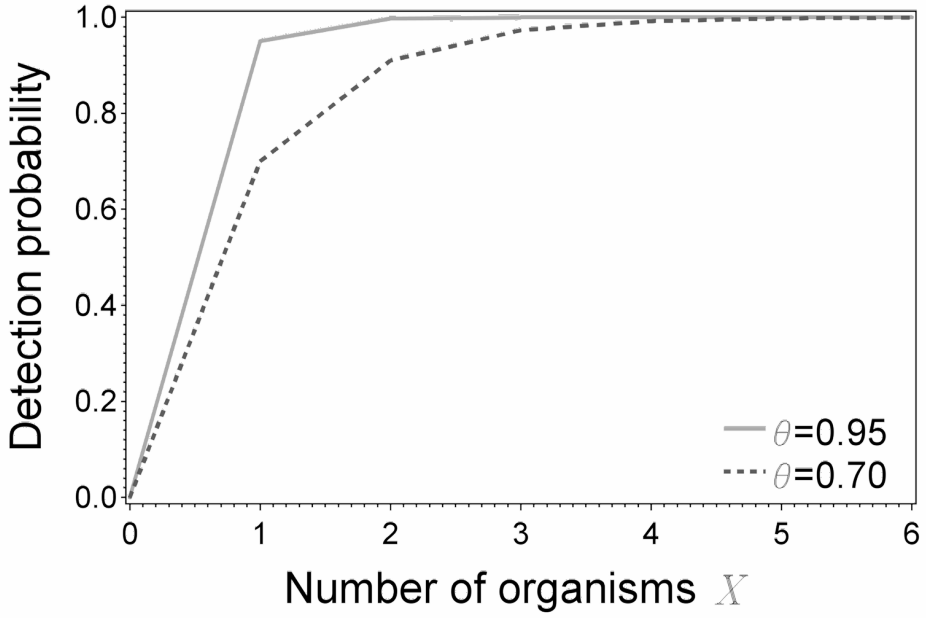
- Compare 2 methods with same λ and consider ratio $\hat{\xi}_A/\hat{\xi}_C$
- Test for non-inferiority: one-sided 95% LCL for $\theta_A/\theta_C > r_0$ or one-sided 95% LCL for $\log(\xi_A) - \log(\xi_C) > \log(r_0)$

Statistical detection model

- Performance cannot be observed directly from results

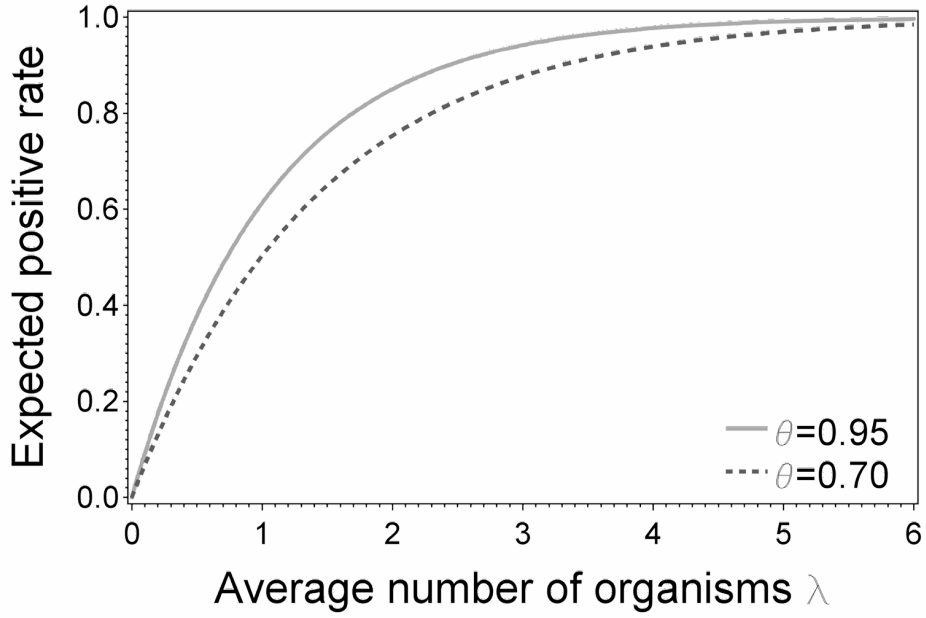
True performance

$$\pi(x) = P(Y = 1|X = x) = 1 - (1 - \theta)^x$$



Observed

$$p = P(Y = 1) = 1 - \exp(-\theta\lambda)$$



➡ Difference in detection proportions is much larger than the observed difference in expected positive rates

Simulations

- Type I error (%) for a single dilution:
  - USP1 (positive rates) vs. gMPN (generalized MPN)
  - 1000 simulations

θ_1, θ_2	Design $n = 200$	λ	$r_0 = 0.8$	
			USP1	gMPN
0.64, 0.8	1x200	0.5	29.0	4.6
		1.0	48.5	5.6
		1.5	68.8	4.5
		2.0	87.8	5.3
		2.5	97.3	4.6
		3.0	99.9	4.7

→ USP1 on positive rates leads to highly inflated Type I error

Simulations

- Power (%) for the 3 approaches:
  - USP1 (positive rates) vs. gMPN for single dilution
  - USP2 (t-test on MPNs) vs. gMPN for multiple dilutions

θ_1, θ_2	Design $n\sim200$	λ	$r_0 = 0.8$			$r_0 = 0.7$		
			USP1	USP2	gMPN	USP1	USP2	gMPN
0.8, 0.8	1x200	0.5	78.1		35.8	93.1		64.8
		1.0	95.5		46.3	99.9		81.4
		1.5	99.5		55.9	100.0		84.0
		2.0	100.0		55.9	100.0		87.2
		2.5	100.0		51.7	100.0		85.6
		3.0	100.0		49.7	100.0		84.0
0.8, 0.8	3x3 x22	4,2,1		37.8	49.7		67.7	82.3
	3x5 x13	4,2,1		44.0	49.2		74.1	82.2

Simulations

- Approach 1 should not be used
  - Concludes non-inferiority too often, even more when spike goes up
- Approach 2 suitable, but can be improved:
  - Single dilution with $\lambda \sim 2$ has optimal power

  If multiple dilutions are used:
  - Choose smaller dilution factor to stay close to the optimum
  - Replace t-test by generalized MPN analysis
    - More power, less risk of failed experiments
  - Use 200 instead of 100 samples, and/or margin 0.7 instead of 0.8
- Generalized MPN approach recommended

Conclusions

- Validation should consider the detection model $\pi(x)$
  - Not recognized in guidelines
- Non-inferiority on positive rates (USP1) incorrect
  - Also affected by spike level
  - False positives could compensate for false negatives
- Non-inferiority on MPN (USP2) correct in binomial model, but not optimal: generalized MPN approach better
- More research needed for other detection models