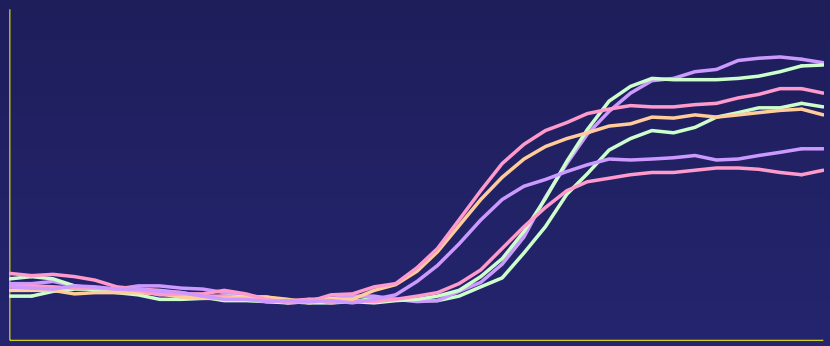
Data Processing in Real Time PCR

Gene expression measurement

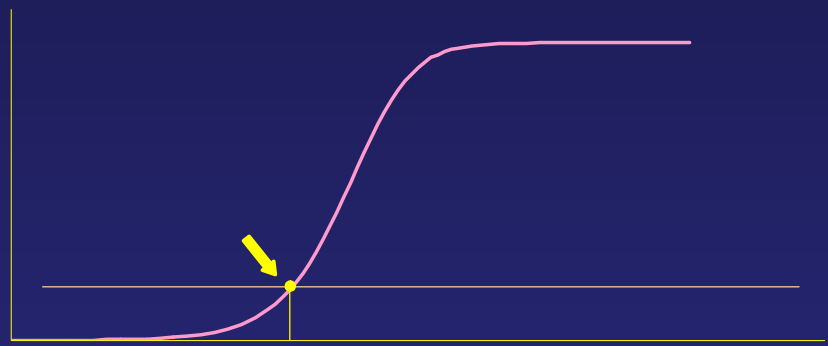
A.A. Larionov, W.R. Miller

Breast Unit Research Group
WGH, Edinburgh, UK

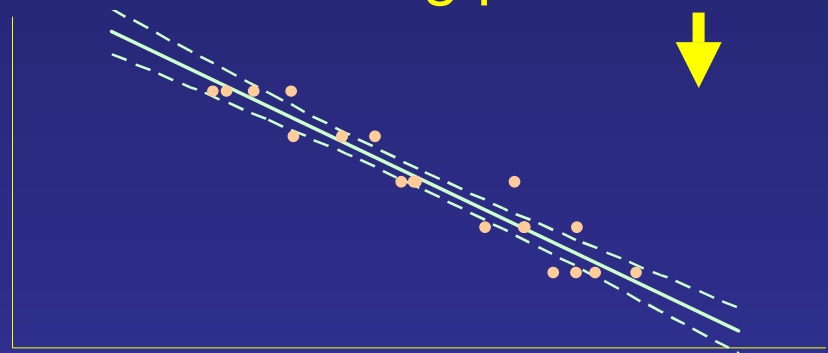
Major Steps in Data Processing



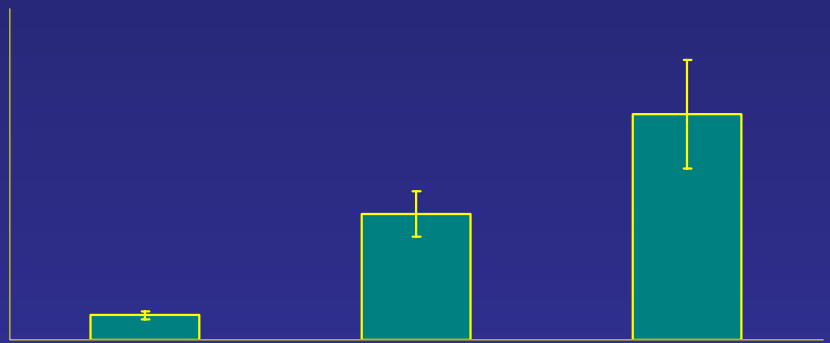
1. Normalization of raw data



2. Determination of crossing points

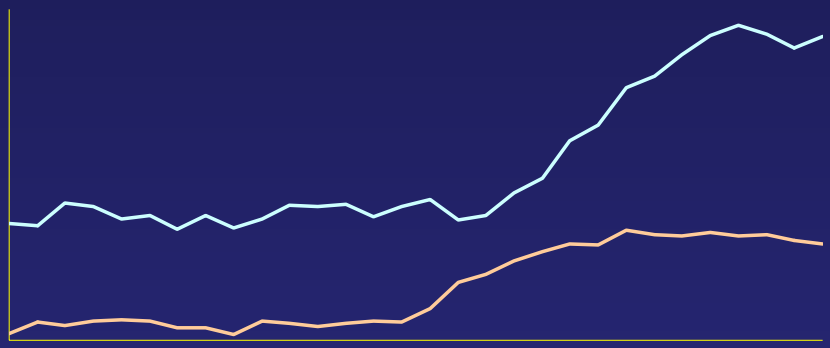


3. Derivation of results from crossing points

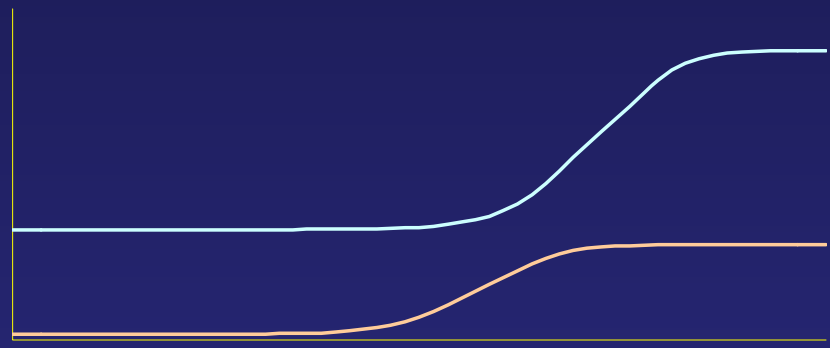


4. Normalization to reference genes

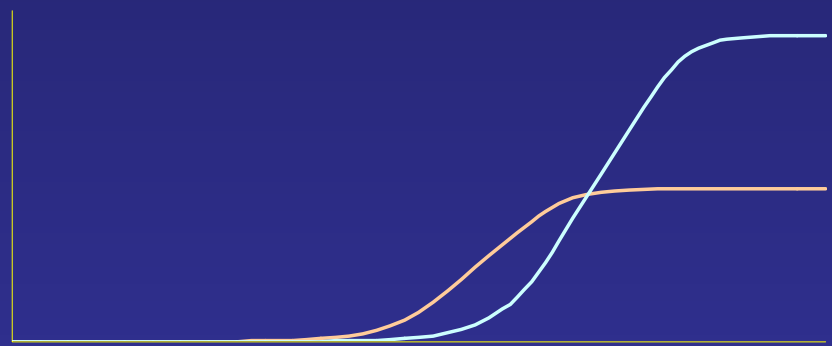
Processing of Raw Measurements



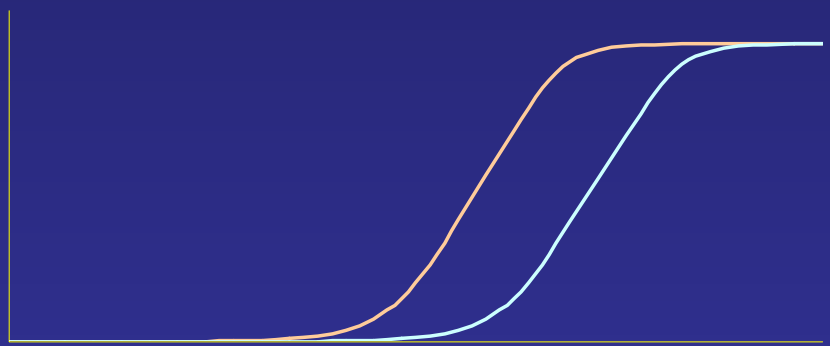
1. Raw readings



2. Smoothing



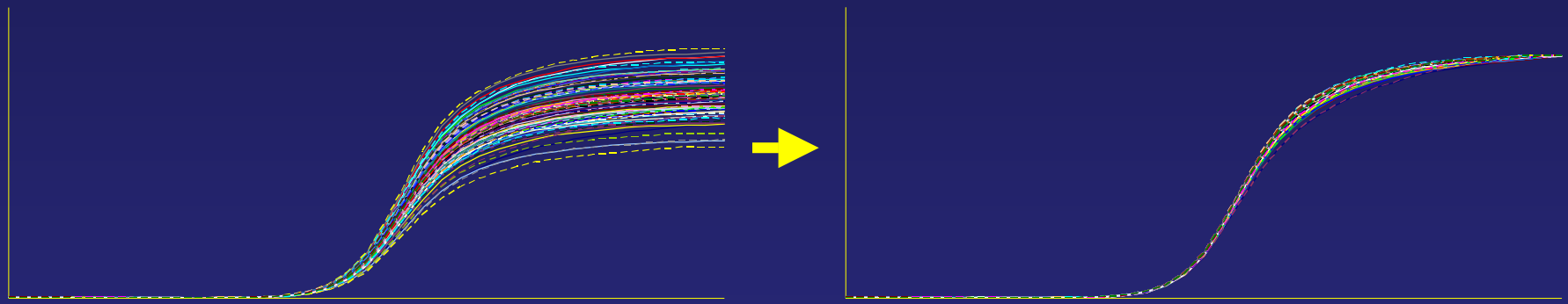
3. Baseline subtraction



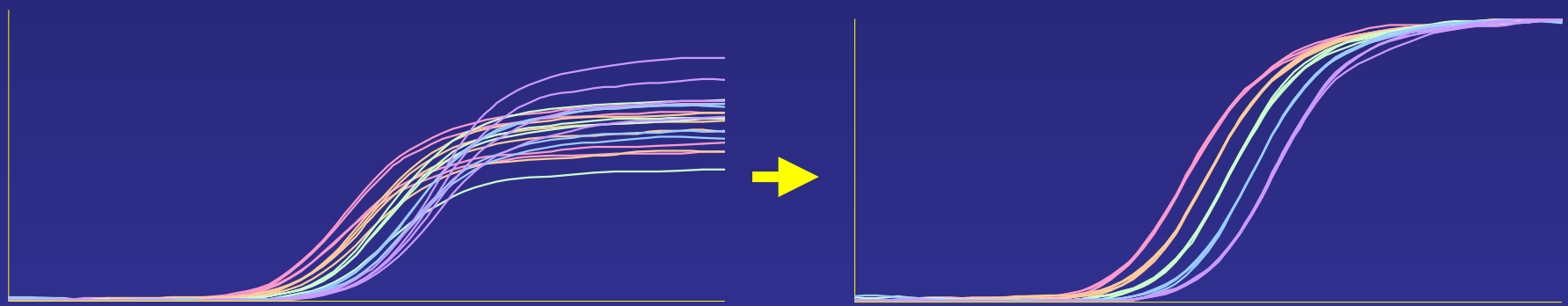
4. Amplitude normalization

Examples of Amplitude Normalization

Replicates (96x)



Serial dilutions (2x)

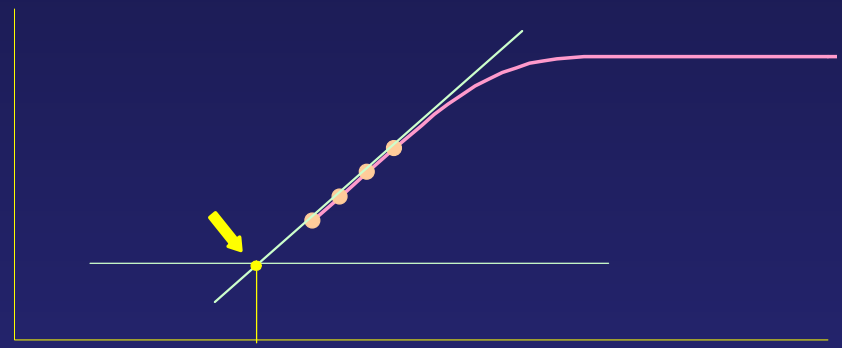


Without amplitude normalization

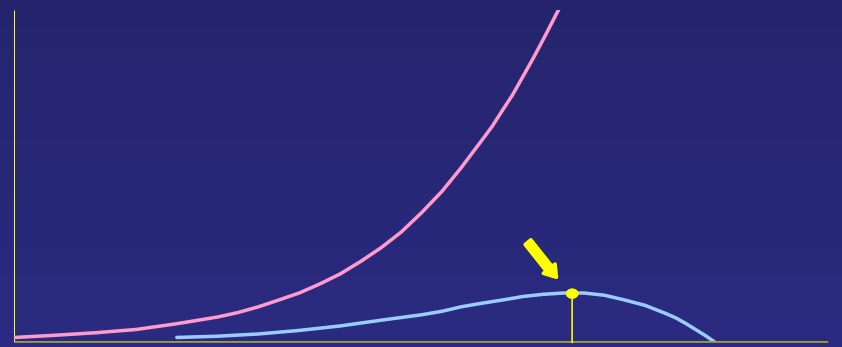
After amplitude normalization

Crossing Points Calculation

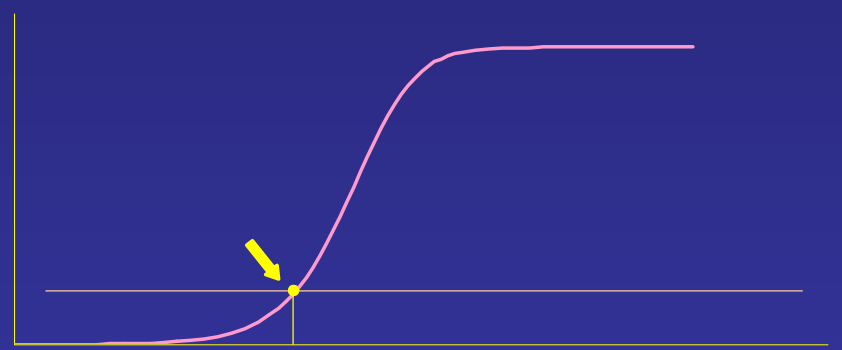
Fit Point method



Second Derivative method



Direct calculation



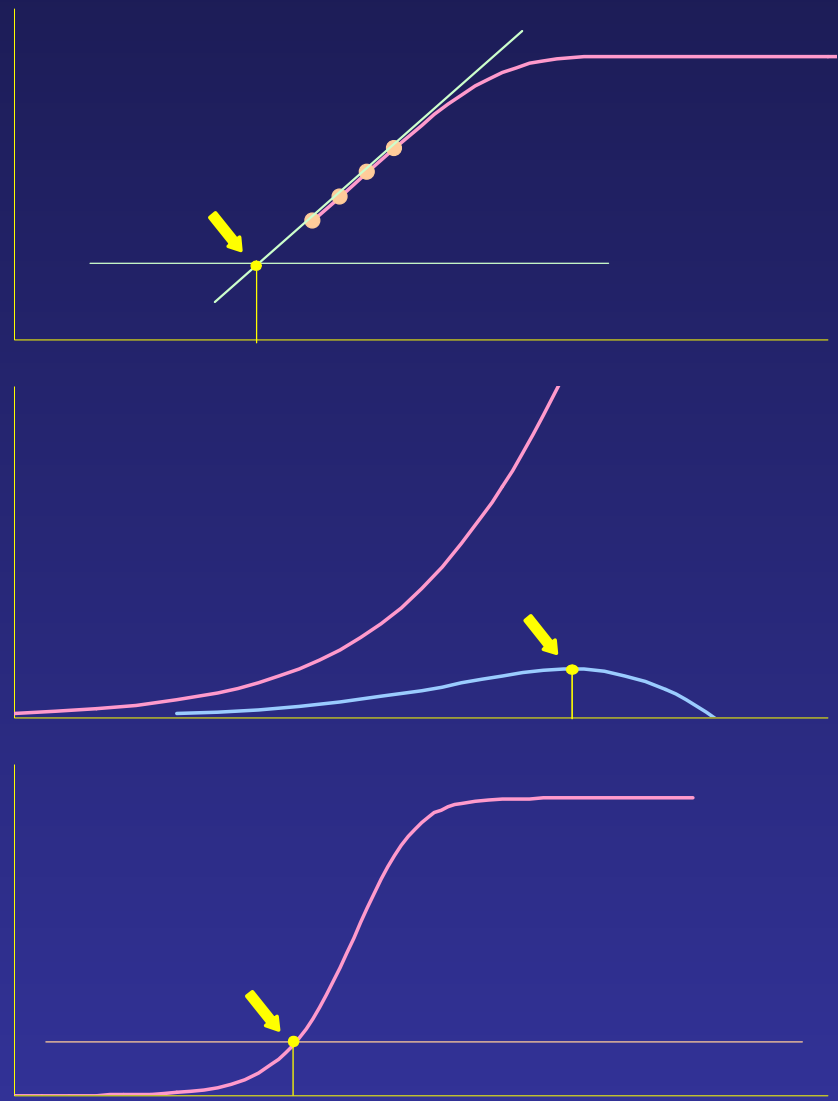
Crossing Points Calculation

Fit Point method

- Minimizes inaccuracies by use of several points
- Log view to stay on exponential phase

Second Derivative method

Direct calculation



Crossing Points Calculation

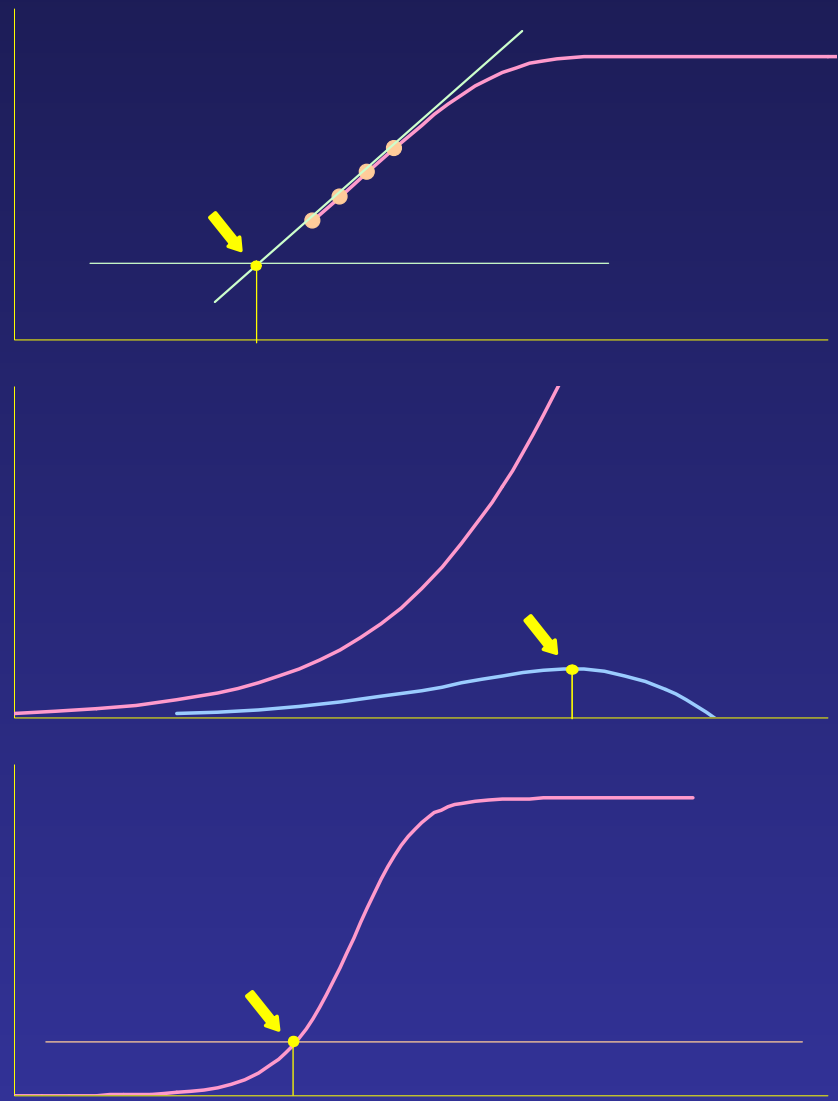
Fit Point method

- Minimizes inaccuracies by use of several points
- Log view to stay on exponential phase

Second Derivative method

- Operator-independent

Direct calculation



Crossing Points Calculation

Fit Point method

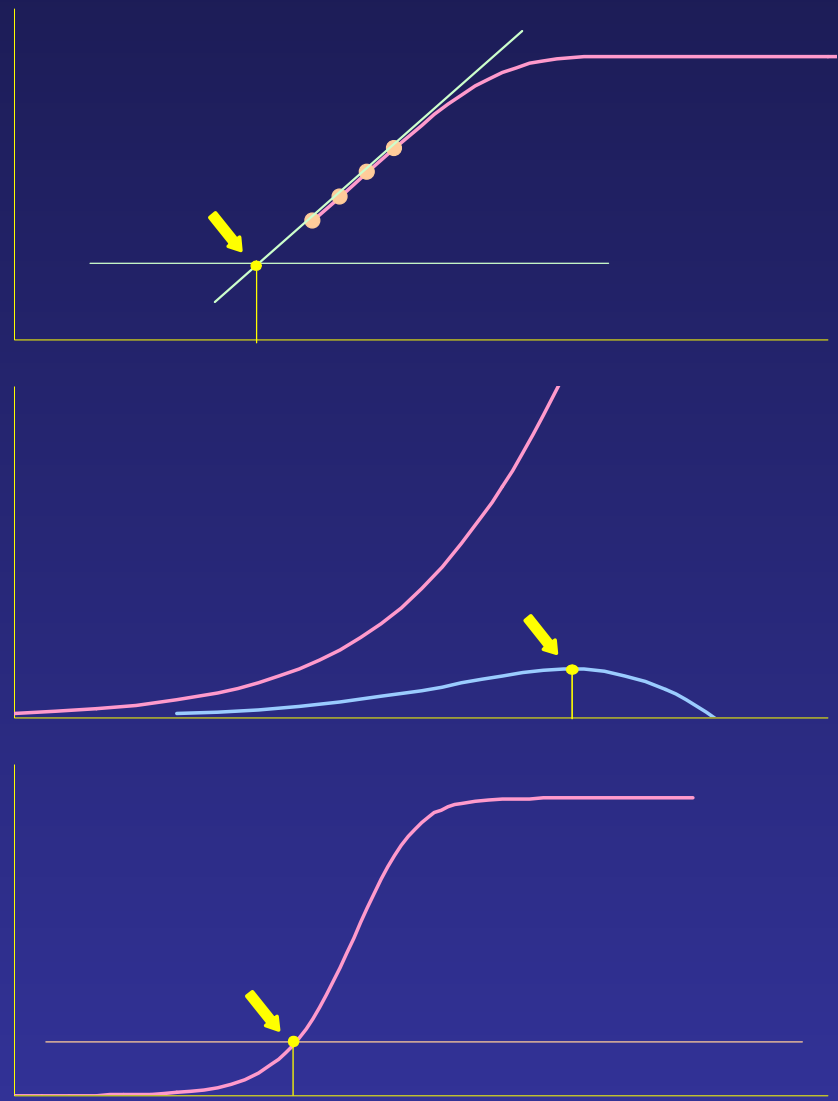
- Minimizes inaccuracies by use of several points
- Log view to stay on exponential phase

Second Derivative method

- Operator-independent

Direct calculation

- Simple mathematical interpretation



Crossing Points Calculation

Fit Point method

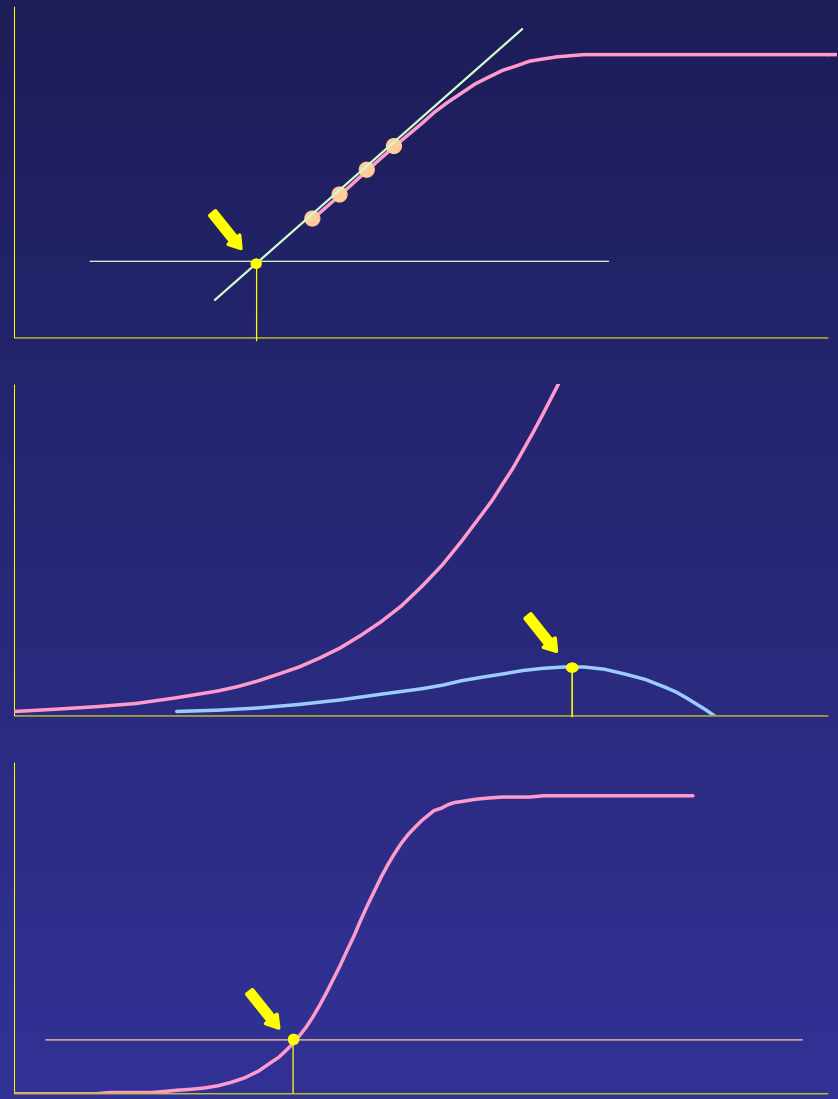
- Minimizes inaccuracies by use of several points
- Log view to stay on exponential phase

Second Derivative method

- Operator-independent

Direct calculation

- Simple mathematical interpretation
- Good results in combination with amplitude normalization and automatic selection of optimal threshold



Derivation of Results from Crossing Points

Best fit values

Errors in replicates

Derivation of Results from Crossing Points

Best fit values

- ***Standard curve method***
  - Simple and Reliable
  - Validated standard curve on each plate
  - Does not need assessment of PCR efficiency
- *Double delta method*
  - Validation is needed
  - Assumptions / assessment of PCR efficiency is needed
  - Less expensive

Errors in replicates

Derivation of Results from Crossing Points

Best fit values

- *Standard curve method*
  - Simple and Reliable
  - Validated standard curve on each plate
  - Does not need assessment of PCR efficiency
- *Double delta method*
  - Validation is needed
  - Assumptions / assessment of PCR efficiency is needed
  - Less expensive

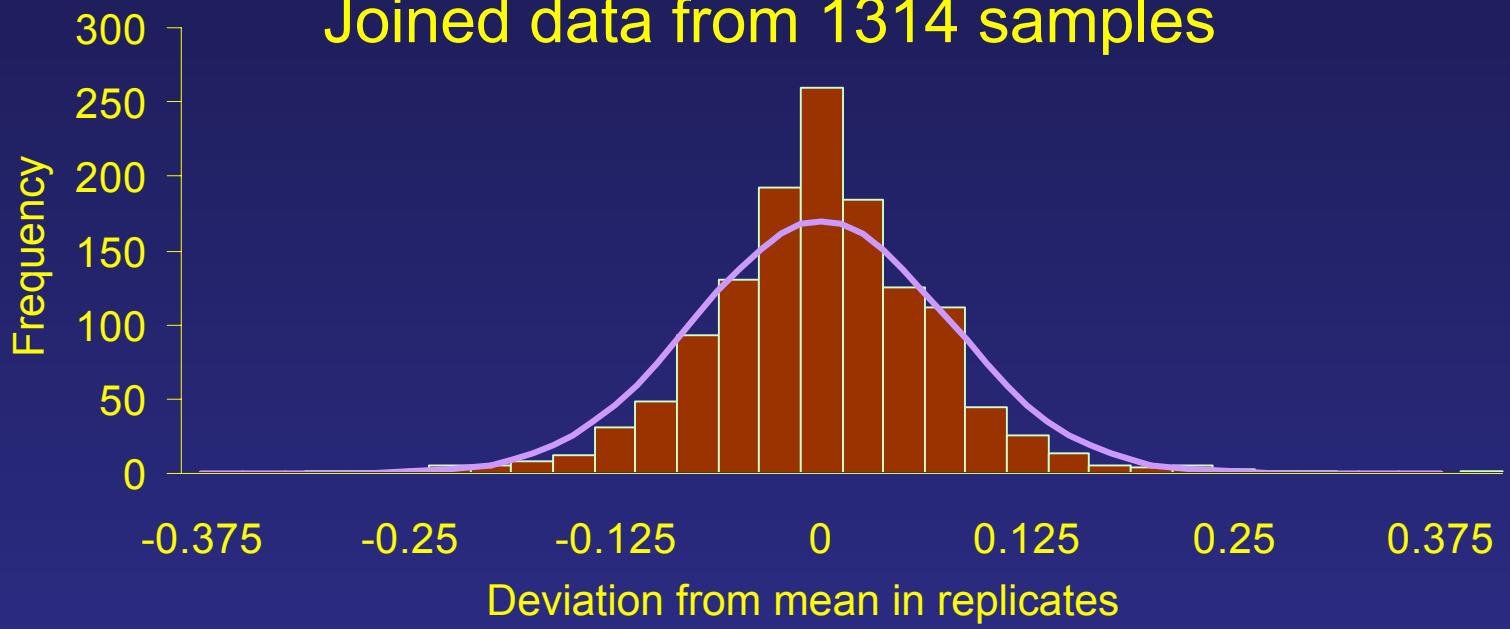
Errors in replicates

- *Parametric methods*
  - Based on Normal distribution
- *Non - Parametric methods*
  - No assumptions regarding error distribution

Crossing Point Distribution in Replicates

Histogram with Normal fit

Joined data from 1314 samples



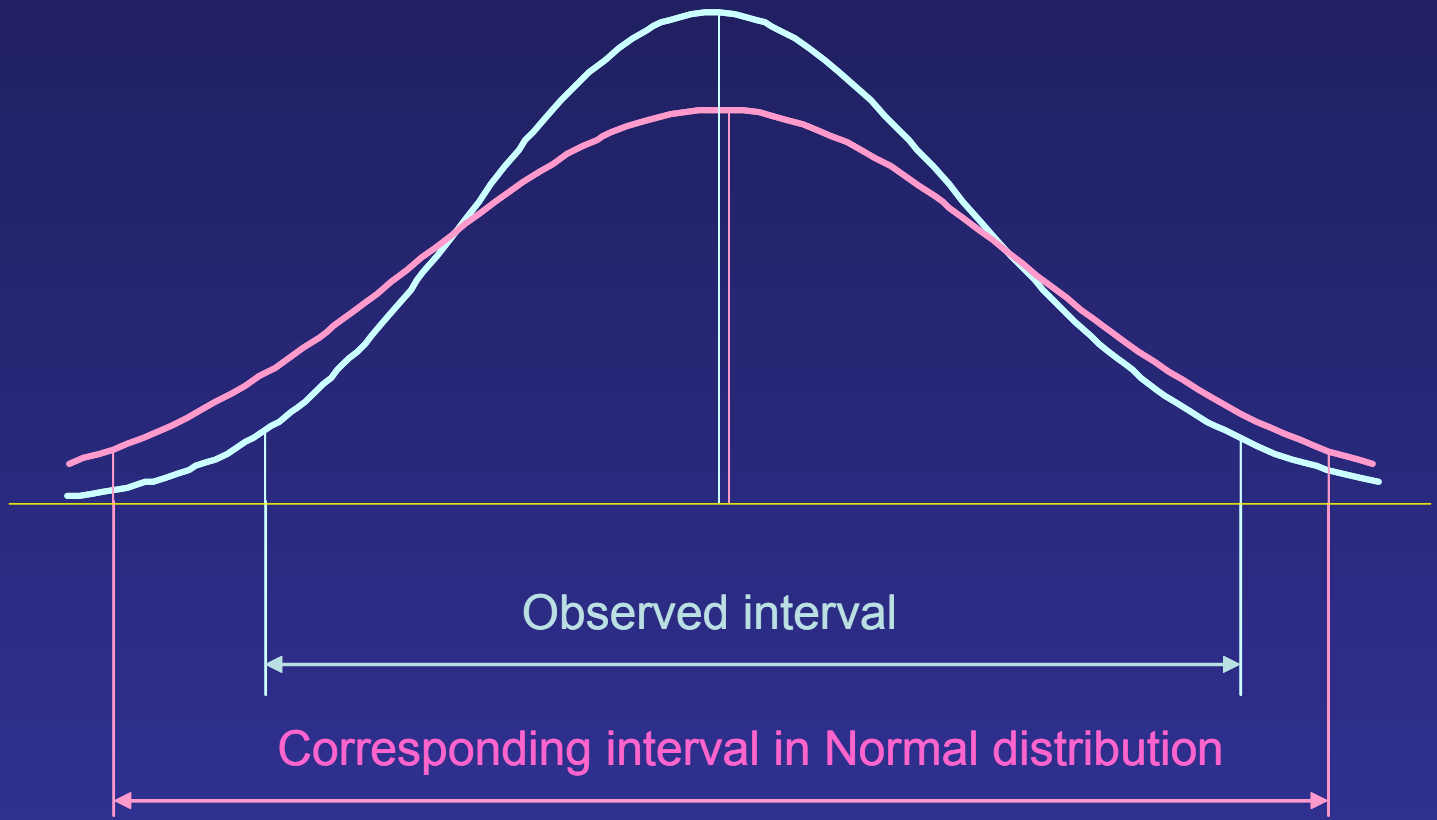
Test for normality

Slightly asymmetric distribution much sharper than Normal

Skewness = 2, Kurtosis = 32

Estimation of Confidence Intervals

Parametric approach can be used
to estimate confidence intervals in Real Time PCR



Tracing of Confidence Intervals

Direct calculation

$$y = 10^{ax+b}$$

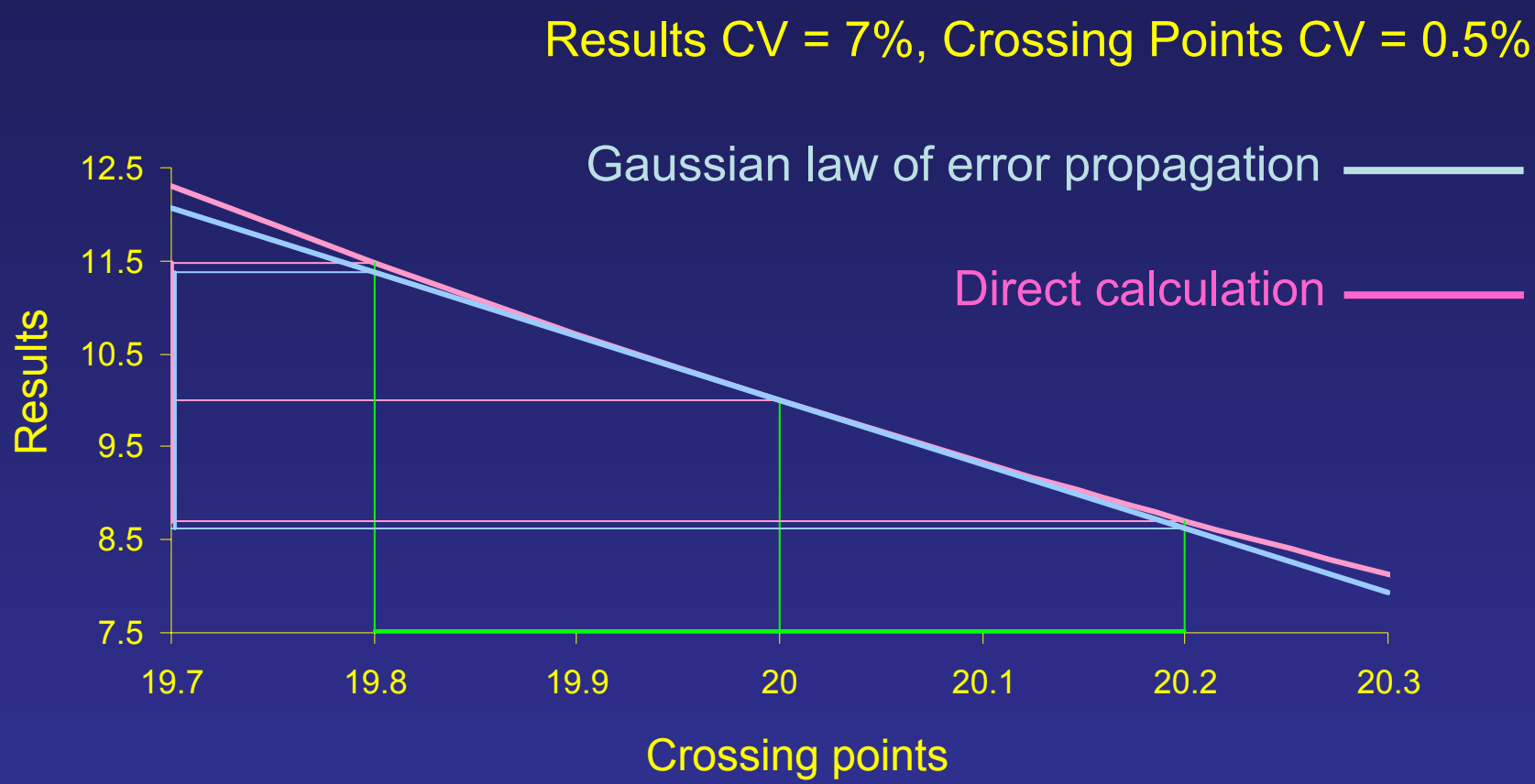
Direct calculation keeps statistical meaning of confidence intervals

Gaussian law of error propagation

$$\Delta y = \frac{dy}{dx}\Delta x$$

Dispersion should not be high:
Crossing point's CV<1% or
Results CV<10%

Example of Error Propagation



Normalizing by Reference Genes

Summarizing of genes

Arithmetic mean

Intuitive and most perceptible

Geometric mean

Has been introduced by Vandesompele *et al, 2002*

Weighted mean

Traces errors

mathematical procedures still might be improved

Normalizing by Reference Genes

Error propagation

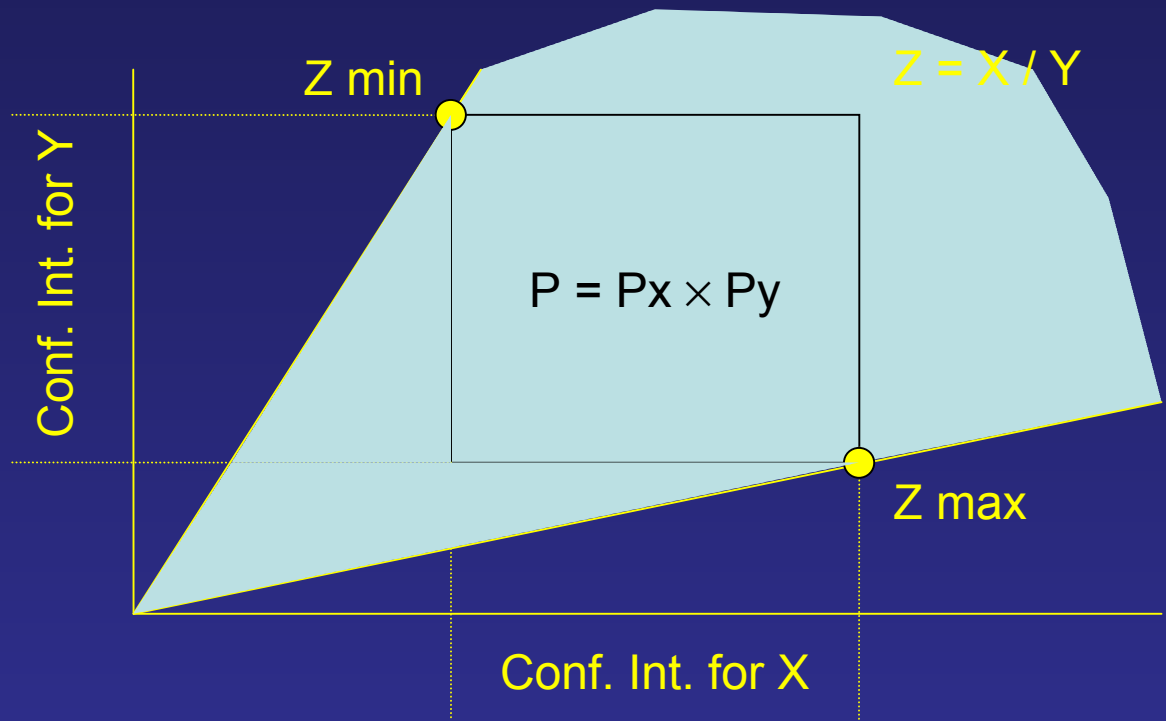
- Simple estimation

Z min = X min / Y max

Z max = X max / Y min

97,5% intervals of X and Y

result to >95% interval for Z



mathematical procedures still might be improved

Summary of Data Processing

Raw measurements

- Smoothing
- Baseline subtraction
- Amplitude normalization

Crossing points

- Direct calculation
- Automatic selection of threshold line

Absolute results

- Standard curves on each plate
- Conf. intervals by Normal Distribution
- Trace errors by law of error propagation

Reference genes

- Weighted mean
- Simplified estimation of conf. intervals

Software Tool

- VBA for MS Excel
- Raw data from CSV files generated by *MJ Opticon Monitor* software
- Results in form of Excel workbook
- Produces charts as well as numeric data
- Keeps all intermediate results available

Concluding Remarks

- ***Amplitude normalization***
  - New step in data processing
- ***Distribution of errors in replicates***
  - Parametric approach for confidence intervals
- ***PCR is not the major source of uncertainty***
  - Sampling, RNA extraction, RT
- ***Remains to be optimized***
  - Mathematics for normalizing by reference genes