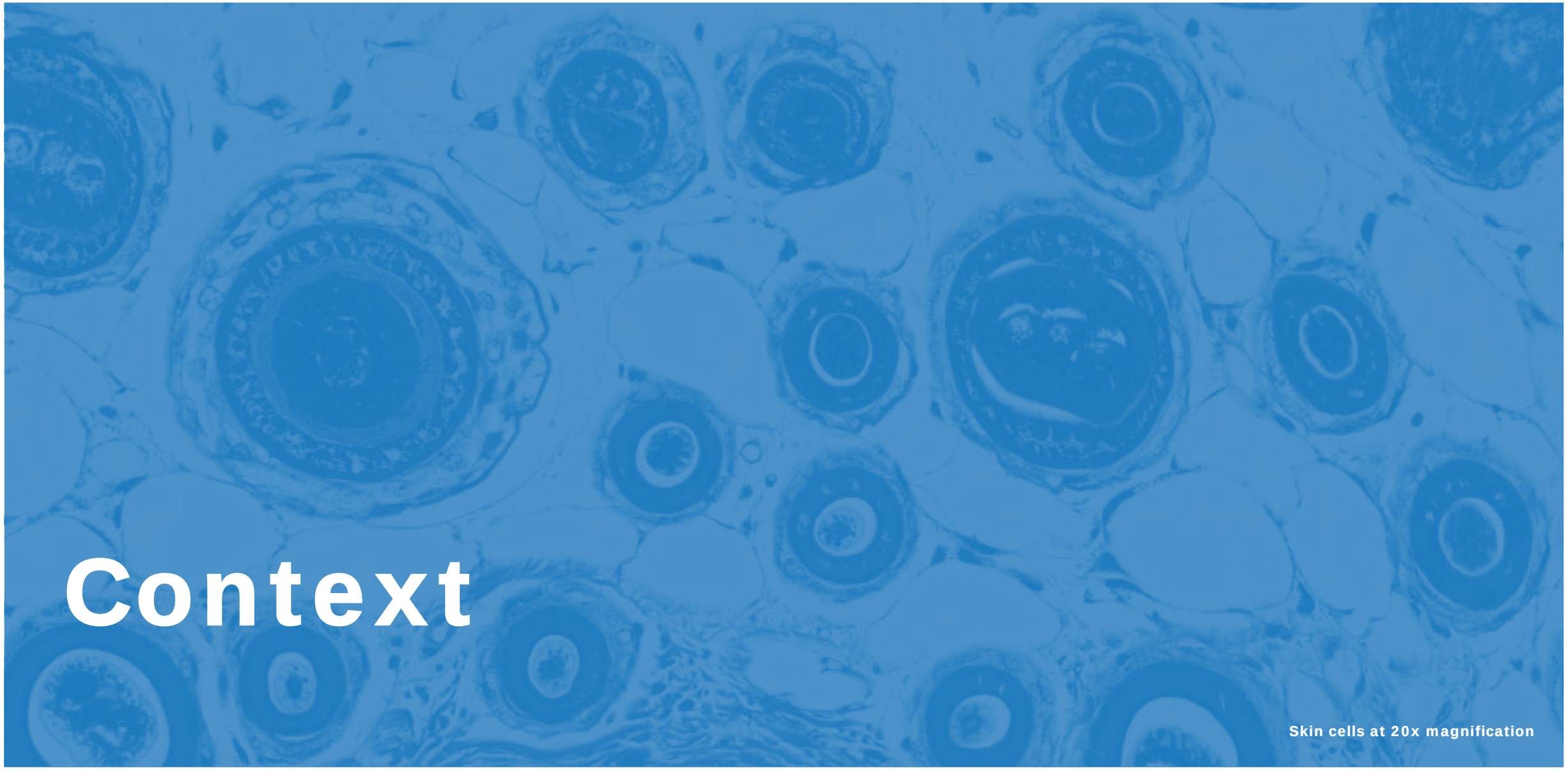
Employing random effects in variance components modelling

Martin Otava
Associate Director
Manufacturing and Applied Statistics, Statistics and Decision Sciences
Janssen Pharmaceutical Companies of Johnson & Johnson
20.10.2022

Context

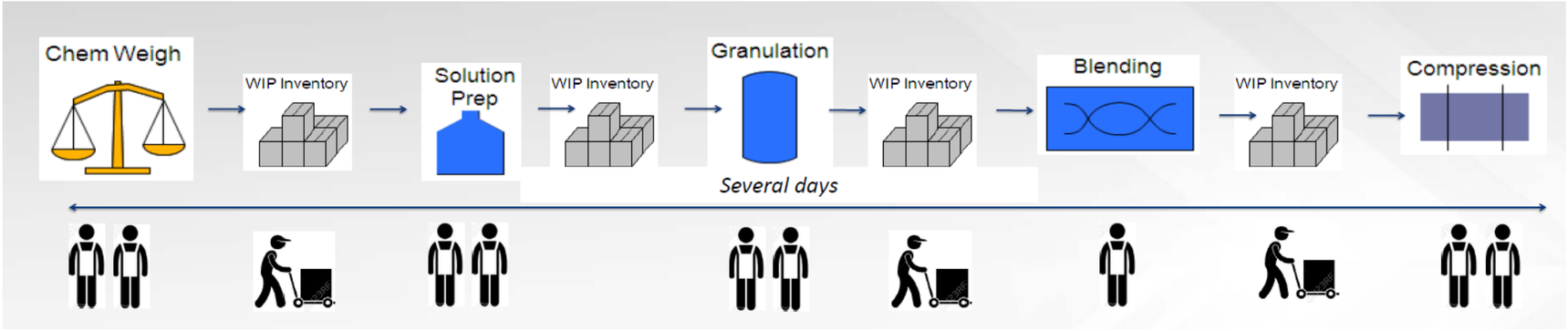


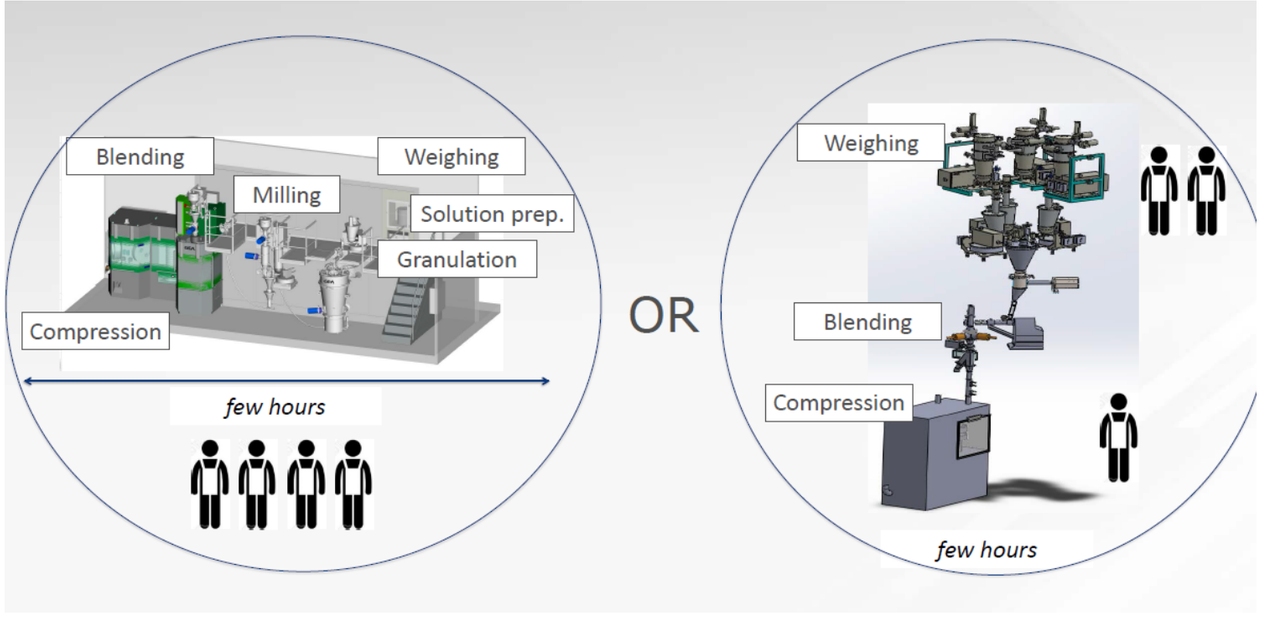
Skin cells at 20x magnification

janssen | PHARMACEUTICAL COMPANIES OF Johnson & Johnson

Continuous manufacture

From batch mode to continuous (CM)





Content & concentration

Important quality attributes

Claim:
20 mg API
200 mg

Content = 100%

Tablet A:
18 mg API
200 mg

Content = 90%

Low concentration

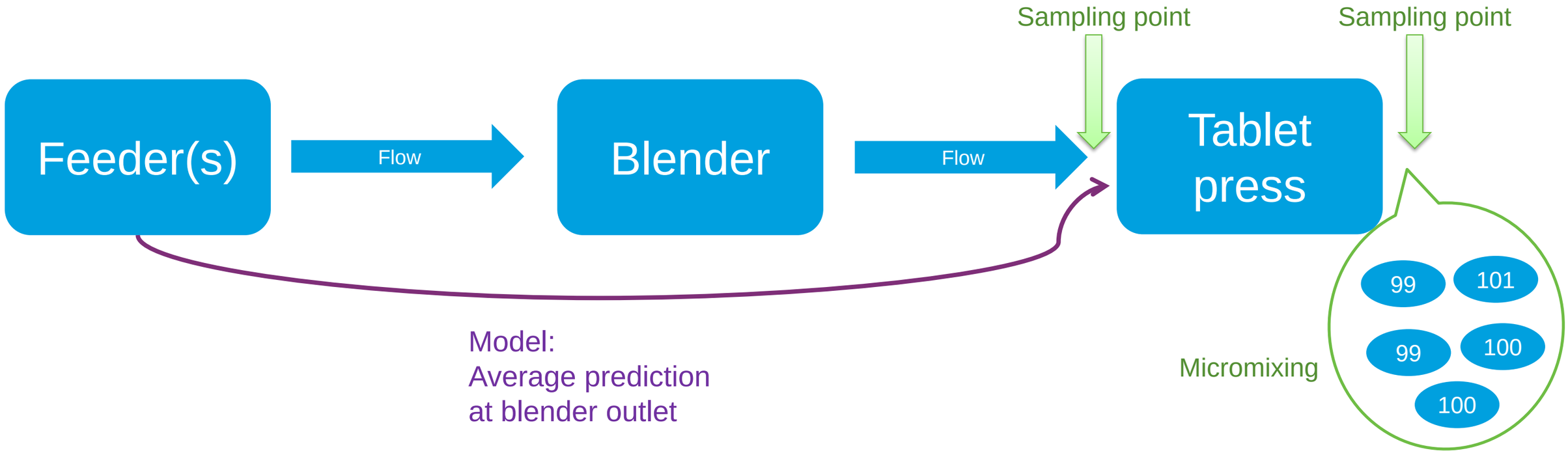
Tablet B:
18 mg API
180 mg

Content = 90%

Low weight

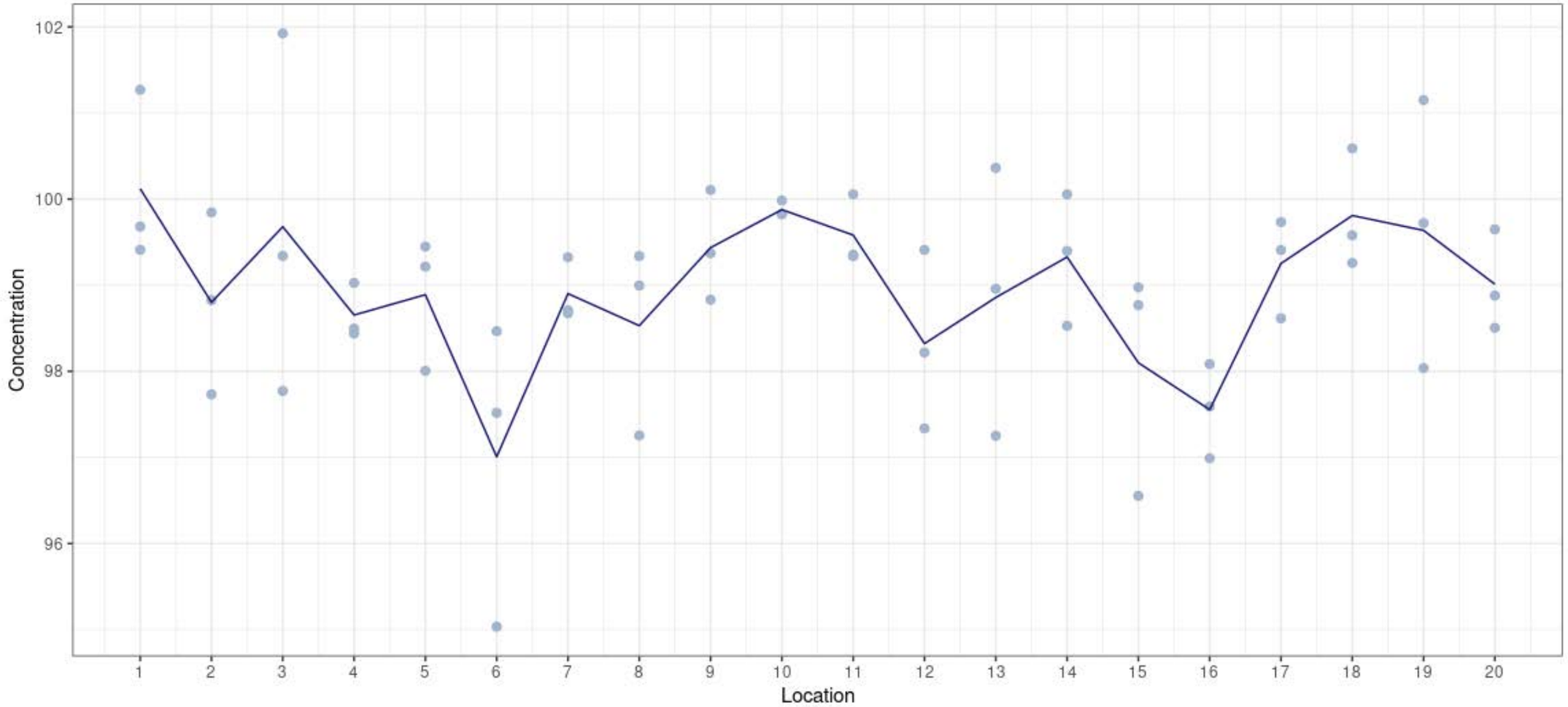
Micro-mixing

Why is it important?



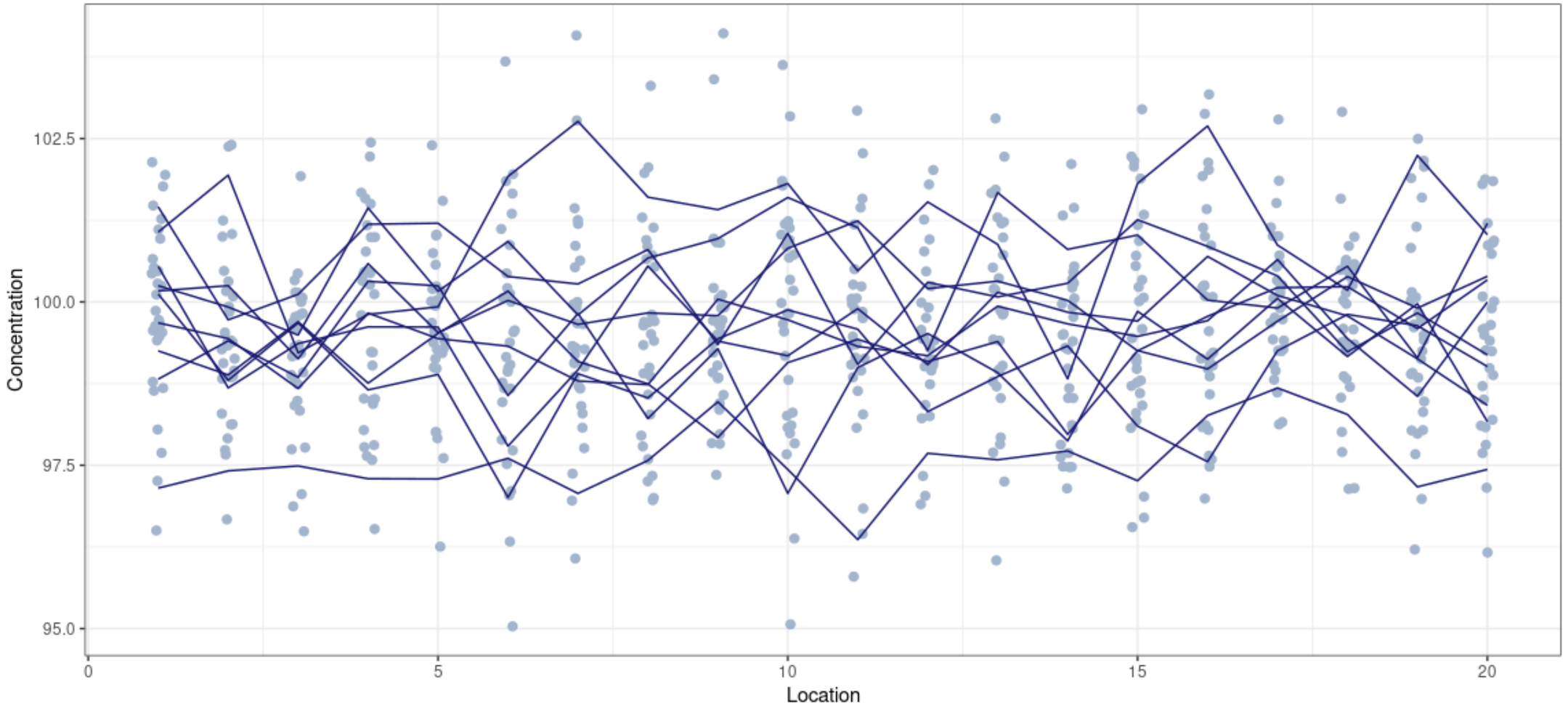
Continuous run [~ batch]

Single run



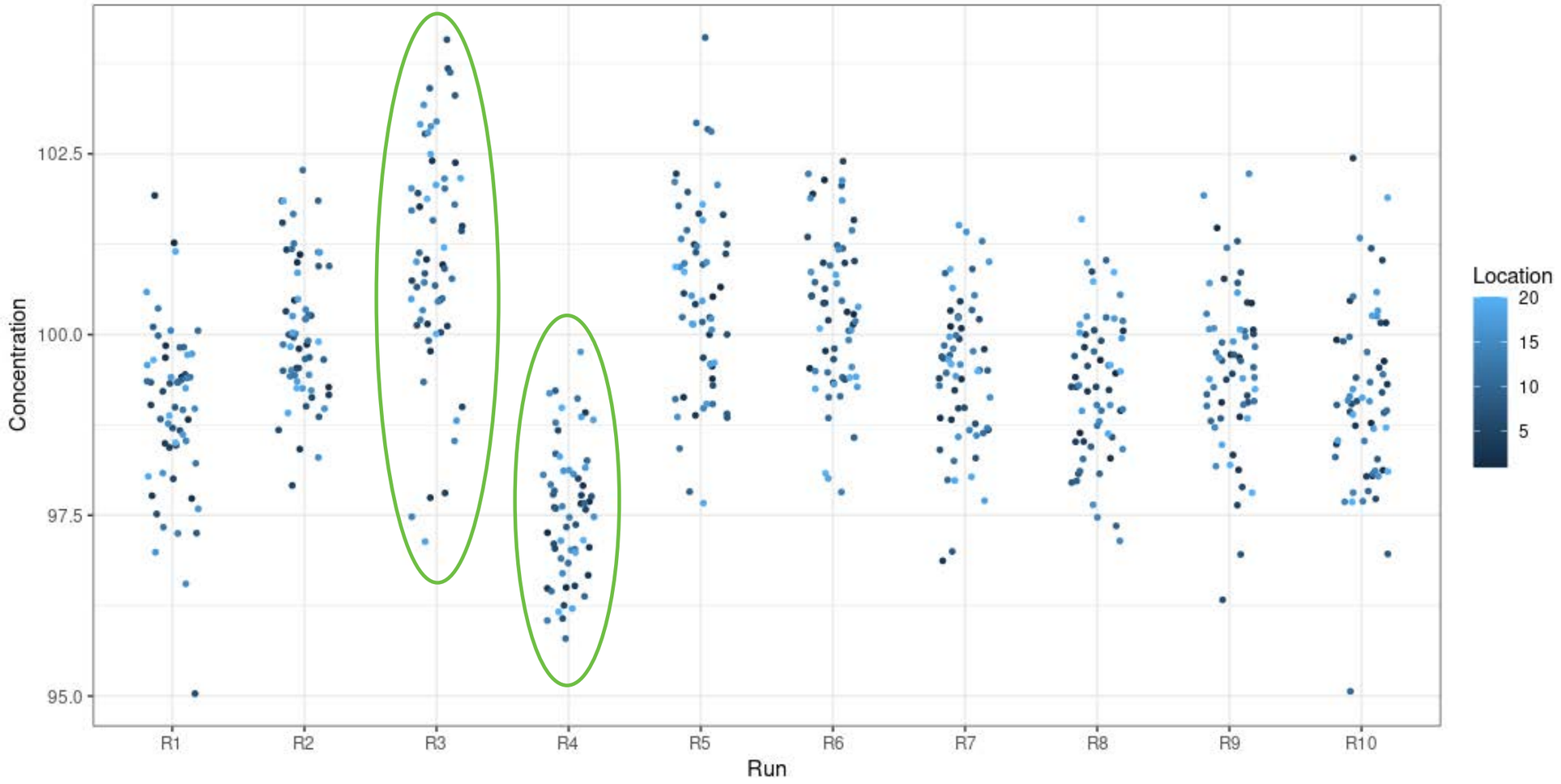
Continuous runs

10 available runs



Continuous run

All runs: different perspective



Model development

Skin cells at 20x magnification

janssen | PHARMACEUTICAL COMPANIES OF Johnson & Johnson

Fixed run effects model

Used in “nearly final process”

Response Location effect

$$Y_{ijk} = R_i + l_{ij} + \varepsilon_{ijk}$$

Run effect Residual error

$$l_{ij} \sim N\left(0, \sigma^2_{loc}\right)$$
$$\varepsilon_{ijk} \sim N\left(0, \sigma^2_{res}\right)$$

Random run effects model

Earlier estimates benefit: more runs

$$Y_{ijk} = \mu + r_i + l_{ij} + \varepsilon_{ijk}$$

$$r_i \sim N(0, \sigma^2_{run})$$
$$l_{ij} \sim N(0, \sigma^2_{loc})$$
$$\varepsilon_{ijk} \sim N(0, \sigma^2_{res})$$

Random run effects model

Earlier estimates risk: not necessarily IID

$$Y_{ijk} = \mu + r_i + l_{ij} + \varepsilon_{ijk}$$

$$r_i \sim N(0, \sigma^2_{run})$$
$$l_{ij} \sim N(0, \sigma^2_{loc})$$
$$\varepsilon_{ijk} \sim N(0, \sigma^2_{res})$$

Modelling variance

Random effect of the run on residuals

$$Y_{ijk} = \mu + r_i + l_{ij} + \varepsilon_{ijk}$$

$$\sigma^2_{res,i} = \sigma^2_{res} + s_i$$

$$r_i \sim N(0, \sigma^2_{run})$$
$$l_{ij} \sim N(0, \sigma^2_{loc})$$
$$\varepsilon_{ijk} \sim N(0, \sigma^2_{res,i})$$

$$s_i \sim N(0, \sigma^2_{runSD})$$

Model R

Modelling variance

Random effect of the run on residuals

$$Y_{ijk} = \mu + r_i + l_{ij} + \varepsilon_{ijk}$$

$$r_i \sim N(0, \sigma^2_{run})$$
$$l_{ij} \sim N(0, \sigma^2_{loc})$$
$$\varepsilon_{ijk} \sim N(0, \sigma^2_{res,i})$$

~~$\sigma^2_{res,i} = \sigma^2_{res} + s_i$~~

$$\log \sqrt{\sigma^2_{res,i}} = \sigma_{res} + s_i$$

$$s_i \sim N(0, \sigma^2_{runSD})$$

Model R

Modelling variance

Random effect of the run on residuals

$$Y_{ijk} = \mu + r_i + l_{ij} + \varepsilon_{ijk}$$

$$r_i \sim N(0, \sigma^2_{run})$$
$$l_{ij} \sim N(0, \sigma^2_{loc,i})$$
$$\varepsilon_{ijk} \sim N(0, \sigma^2_{res,i})$$

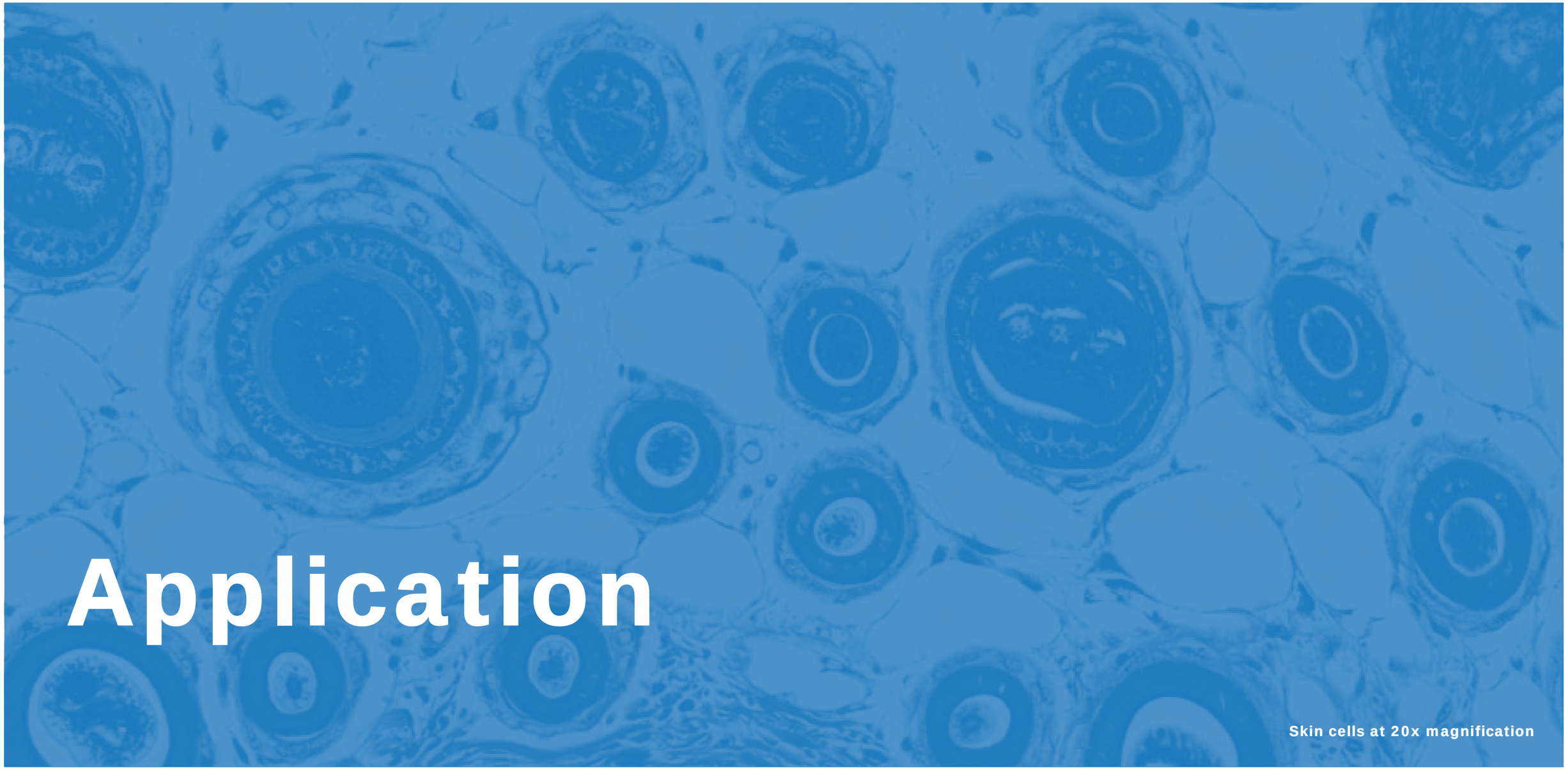
$$\log \sqrt{\sigma^2_{res,i}} = \sigma_{res} + s_i$$

$$\sigma^2_{loc,i} = \sigma_{loc} + t_i$$

$$s_i \sim N(0, \sigma^2_{runSD})$$
$$t_i \sim N(0, \sigma^2_{runLocSD})$$

Model RL

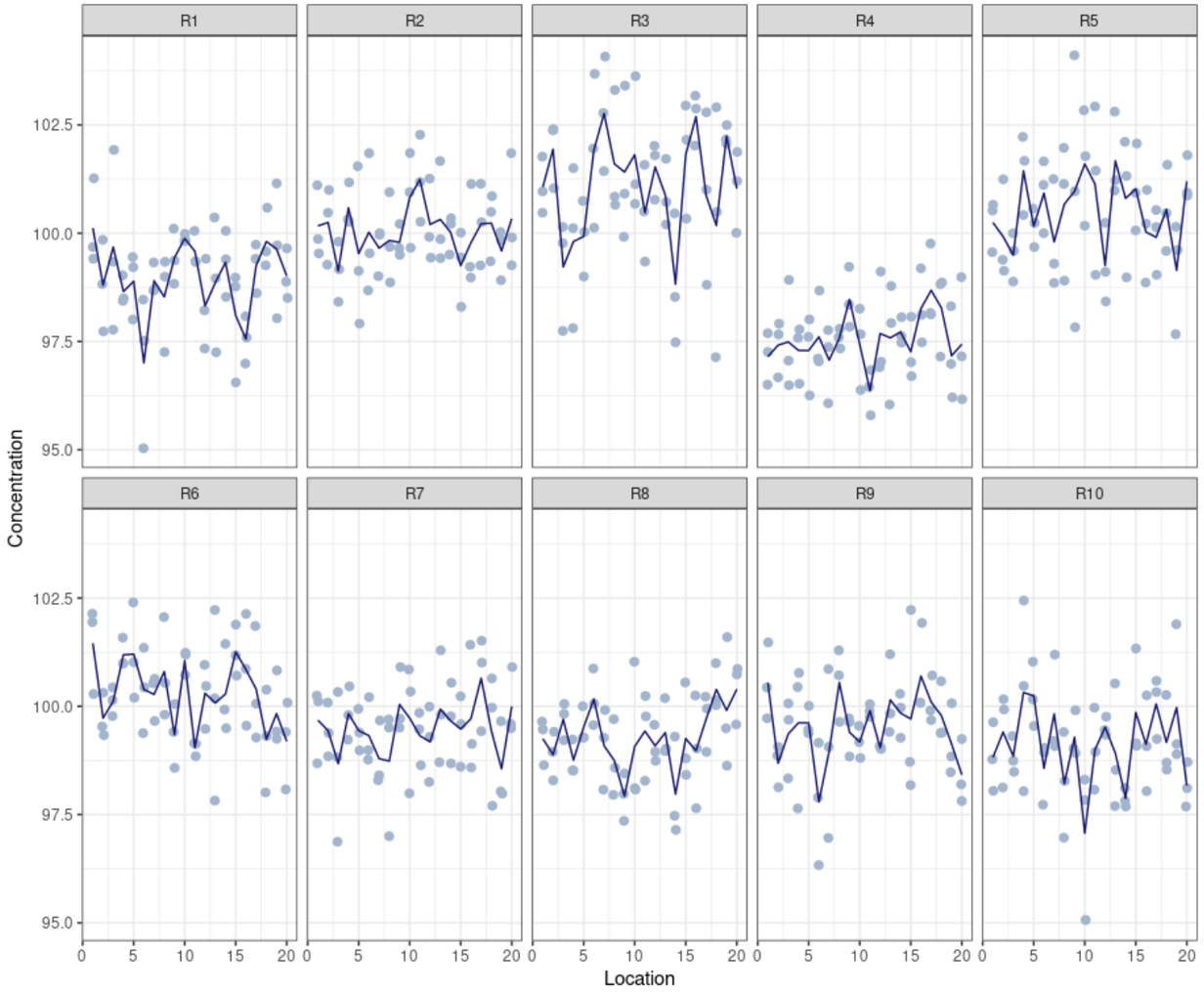
Application



Skin cells at 20x magnification

Case study

Artificial data resembling real CM line data



Implementation

R code in brms

Model R

```
modelR <-
brm(bf(CU ~ 1 +
                (1|Run) +
                (1|Run_Loc),
       sigma ~ 1 + (1|Run)
       ),
       data = dataCUsim,
       chains = 3, iter = 5000, warmup= 1000,
       thin= 1, seed = 1235, cores= 3,
       control = list(     adapt_delta = 0.95,
                           max_treedepth = 10)
)
```

Model RL

```
modelRL <-
brm(bf(CU ~ 1 +
                (1|Run) +
                (1|gr(Run_Loc, by = Run)),
       sigma ~ 1 + (1|Run)
       ),
       data = dataCUsim,
       chains = 3, iter = 5000, warmup= 1000,
       thin= 1, seed = 1235, cores= 3,
       control = list(     adapt_delta = 0.95,
                           max_treedepth = 10)
)
```

Implementation

R code in brms

Model R

```
modelR <-
brm(bf(CU ~ 1 +
                (1|Run) +
                (1|Run_Loc),
       sigma ~ 1 + (1|Run)
       ),
       ...
)
summary(modelR)
```

```
Population-Level Effects:
                Estimate Est.Error l-95% CI u-95% CI
Intercept          99.58      0.36    98.87   100.32
sigma_Intercept     0.08      0.07    -0.05     0.21

Group-Level Effects:
~Run (Number of levels: 10)
                  Estimate Est.Error l-95% CI u-95% CI
sd(Intercept)         1.11      0.31     0.68     1.88
sd(sigma_Intercept)   0.16      0.06     0.06     0.31
```

```
chainsR <- as_draws_matrix(modelR)

mean(exp(chainsR[, "b_sigma_Intercept"]))
quantile(exp(chainsR[, "b_sigma_Intercept"]), c(0.025, 0.975))
```

Model RL

```
modelRL <-
brm(bf(CU ~ 1 +
                (1|Run) +
                (1|gr(Run_Loc, by = Run)),
       sigma ~ 1 + (1|Run)
       ),
       ...
)
summary(modelLR)
```

```
~Run_Loc (Number of levels: 200)
                   Estimate Est.Error l-95% CI u-95% CI
sd(Intercept:RunR1)     0.49      0.24     0.05     0.98
sd(Intercept:RunR2)     0.19      0.14     0.01     0.54
sd(Intercept:RunR3)     0.85      0.29     0.27     1.46
sd(Intercept:RunR4)     0.22      0.16     0.01     0.58
sd(Intercept:RunR5)     0.38      0.23     0.02     0.89
sd(Intercept:RunR6)     0.43      0.22     0.03     0.90
sd(Intercept:RunR7)     0.20      0.15     0.01     0.56
sd(Intercept:RunR8)     0.35      0.21     0.02     0.78
sd(Intercept:RunR9)     0.41      0.23     0.03     0.87
sd(Intercept:RunR10)    0.56      0.25     0.08     1.08
```

```
chainsRL <- as_draws_matrix(modelRL)
apply(chainsRL[, paste0("sd_Run_Loc__Intercept:RunR", 1:10)], 1, mean)
apply(chainsRL[, paste0("sd_Run_Loc__Intercept:RunR", 1:10)], 1, sd)
```

Case study results

Model with random residual error across batches (Model R)

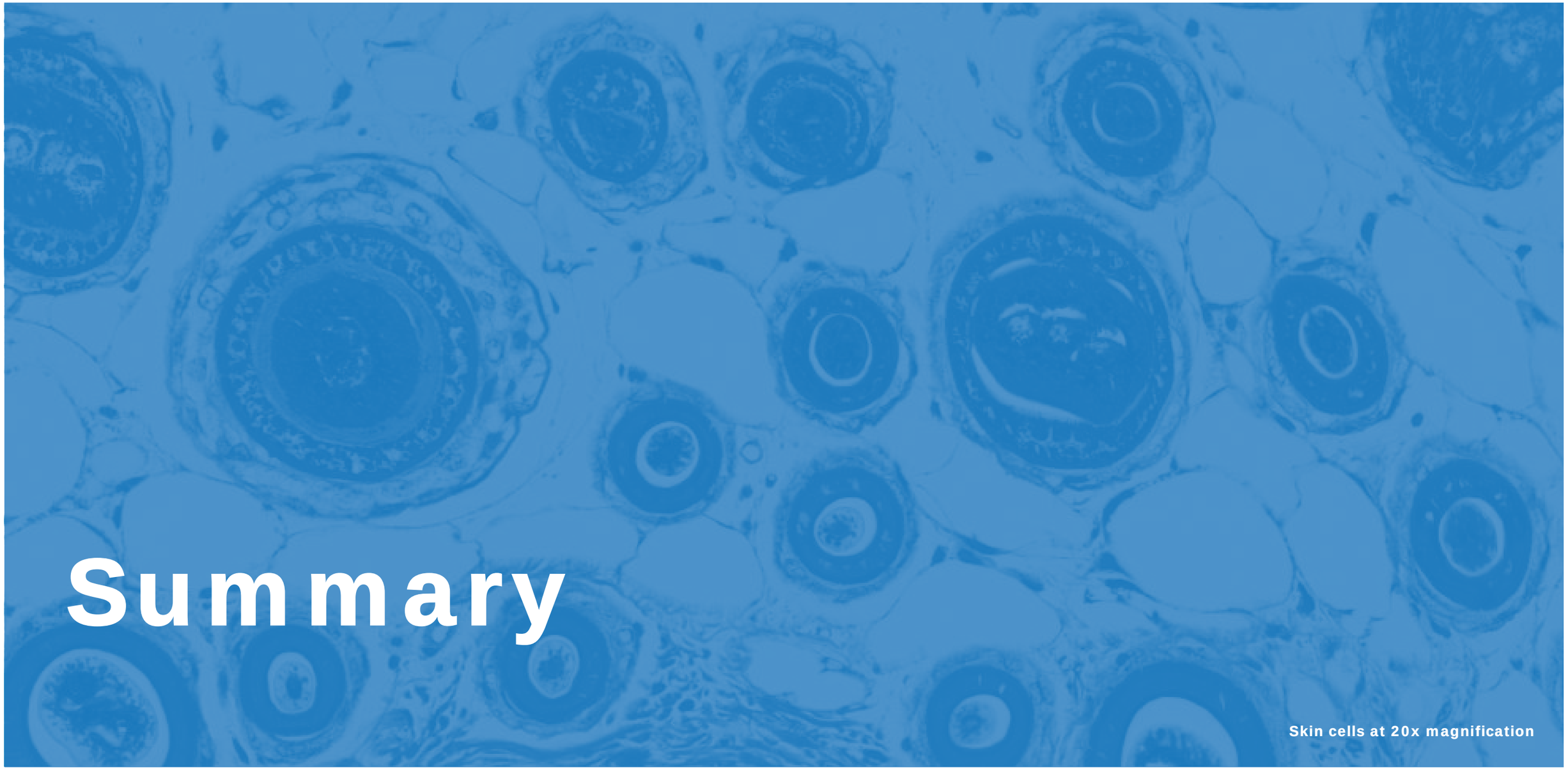
Parameter	Interpretation	Estimate	95% Credible limit
μ	Global process mean	99.58	(98.87, 100.32)
σ_{run}	Run SD	1.11	(0.68, 1.88)
σ_{loc}	Location SD	0.32	(0.08, 0.50)
σ_{res}	Log Average/Population residual SD	0.08	(-0.05, 0.21)
$\mathbf{exp}(\boldsymbol{\sigma_{res}})$	**Average/Population residual SD**	**1.09**	**(0.96, 1.24)**
σ_{runSD}	Run SD on Log Residual SD	0.16	(0.06, 0.31)
NA	Run SD on Residual SD	simulation	simulation

Case study results

Model with random residual and location error across batches (Model RL)

Parameter	Interpretation	Estimate	95% Credible limit
μ	Global process mean	99.59	(98.88, 100.30)
σ_{run}	Run SD	1.10	(0.67, 0.1.85)
$\boldsymbol{\sigma_{loc}}$	**Average/Population Location SD**	**0.41**	**(0.28, 0.55)**
$\boldsymbol{\sigma_{runLocSD}}$	**Run SD on Population Location SD**	**0.29**	**(0.17, 0.44)**
σ_{res}	Log Average/Population residual SD	0.07	(-0.04, 0.18)
$\exp(\sigma_{res})$	Average/Population residual SD	1.07	(0.96, 1.19)
σ_{runSD}	Run SD on Log Population Residual SD	0.12	(0.02, 0.27)
NA	Run SD on Population Residual SD	simulation	simulation

Summary



Skin cells at 20x magnification

janssen | PHARMACEUTICAL COMPANIES OF Johnson & Johnson

Key takeaway

- Random effects on variance are straightforward!
  - Residual error variation: different degree of micromixing
  - Location SD variation: different stability of the run

- Batch mode applicability

- Extension: multiple measurements of the same sample => analytical and true product variability

Results disclaimer: Not a real data, only simulation!

Thank you for your attention!


janssen
PHARMACEUTICAL COMPANIES OF
Johnson&Johnson