

# STATISTICAL POPULATION MODELING AND ANALYSIS OF PK DATA

## An Introduction

---

Marie Davidian  
Department of Statistics  
North Carolina State University



<http://www.stat.ncsu.edu/~davidian>

# Outline

---

1. Introduction and objectives
2. Pharmacokinetic models
3. Population PK modeling and analysis
4. Hierarchical statistical model formulation
5. Variability and uncertainty
6. Parameter estimation
7. Concluding remarks

# Introduction and Objectives

---

**Pharmacokinetic (PK) models:** *Mathematical models* for *pharmacokinetic* mechanisms taking place within *physiological systems*

- Characterize *concentrations over time* within the system . . .
- . . . in terms of *parameters* describing processes of *A*bsorption, *D*istribution, *M*etabolism, *E*limination

**PK models in risk assessment:** Of *increasing interest* as a basis for

- Characterizing *time course* of concentrations of *hazardous agents* in different parts of the body under known *exposure* levels/patterns
- *Extrapolation* from animals to humans, high to low exposure levels
- Relationships between *delivered dose* (e.g., in *target tissues*) and *adverse response*

# Introduction and Objectives

---

## Key messages:

- Such analyses are based on *fitting* PK models to *data* (historical and experimental)
- Must be done within a *statistical framework* that acknowledges the *sources of variability* in data . . .
- . . . and gives a basis for *quantification of uncertainty* associated with fitting the model and further analyses based on it
- “*Population PK modeling and analysis*”

# Introduction and Objectives

---

## Objectives:

- Describe the perspective and goals of *population analysis*
- Describe the *statistical model framework*
- Explain how it represents *variability* and provides a basis for assessment of *uncertainty*, and why this is *essential*
- Review popular *estimation methods* for fitting the model
- Raise important topics for *discussion*

# Pharmacokinetic Models

---

## PK models:

- Represent the body as a *system of compartments* . . .
- . . . leading to *deterministic system of equations* whose solution yields *expressions for concentrations* in each compartment
- Used for *decades* to study *time course* and *fate* of *parent* or *metabolite* concentrations of *pharmaceutical agents* in humans to develop *dosing regimens* with *great success*
- Typically grossly simple, *empirical* models (one, two compartments) that involve *just a few PK parameters*, yield *explicit* compartment-specific concentration expressions

# Pharmacokinetic Models

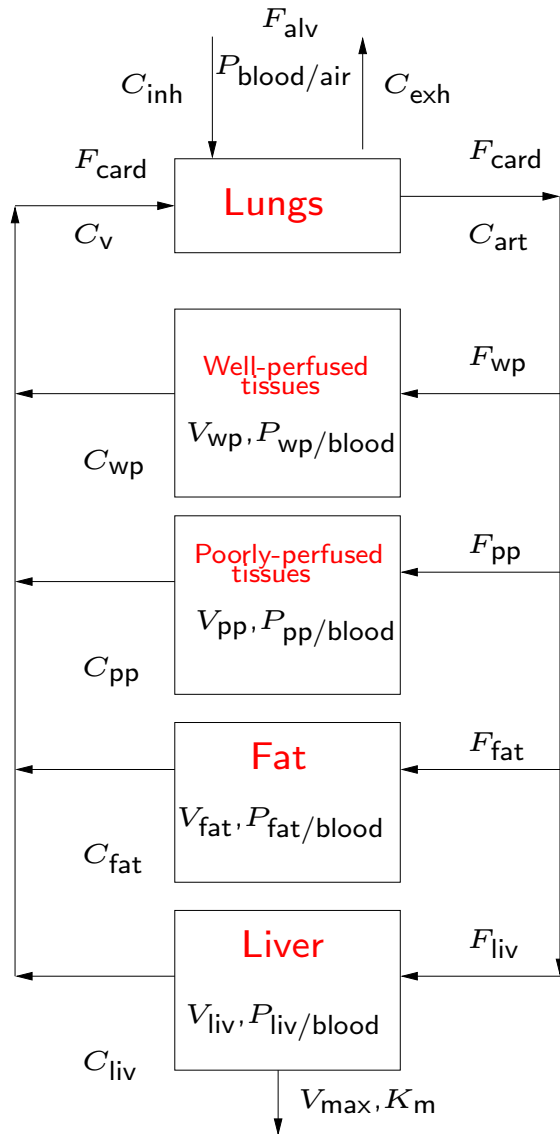
---

## Physiologically Based Pharmacokinetic Models (PBPK) :

- A “*more realistic*” description of the body (lots of compartments)
- Explicit expressions for compartment-specific concentrations *not possible* in general (must be obtained *numerically*)
- *Lots* of PK parameters
  - Compartment volumes  $V$
  - Partition coefficients  $P$
  - Flow rates  $F$
  - Michaelis-Menten metabolic clearance  $V_{\max}$ ,  $K_m$
  - Some *known or measurable*
  - Others *unknown*

# Pharmacokinetic Models

## Common PBPK model for VOCs and form of equations:



$$C_{art} = \frac{F_{card}C_{ven} + F_{alv}C_{inh}}{F_{card} + F_{alv}/P_{blood/air}}, \quad C_{ven} = \sum_s \frac{F_s C_s}{F_{card}}$$

$$C_{exh} = (1 - \delta) \frac{C_{art}}{P_{blood/air}} + \delta C_{inh}$$

$$\frac{dC_s}{dt} = \frac{F_s}{V_s} \left( C_{art} - \frac{C_s}{P_{s/blood}} \right), \quad s = wp, pp, fat$$

$$\frac{dC_{liv}}{dt} = \frac{F_{liv}}{V_{liv}} \left( C_{art} - \frac{C_{liv}}{P_{liv/blood}} \right) - R_{liv} \quad (s = liv),$$

$$R_{liv} = \frac{V_{max}C_{liv}}{V_{liv}(K_m + C_{liv})},$$

# Pharmacokinetic Models

---

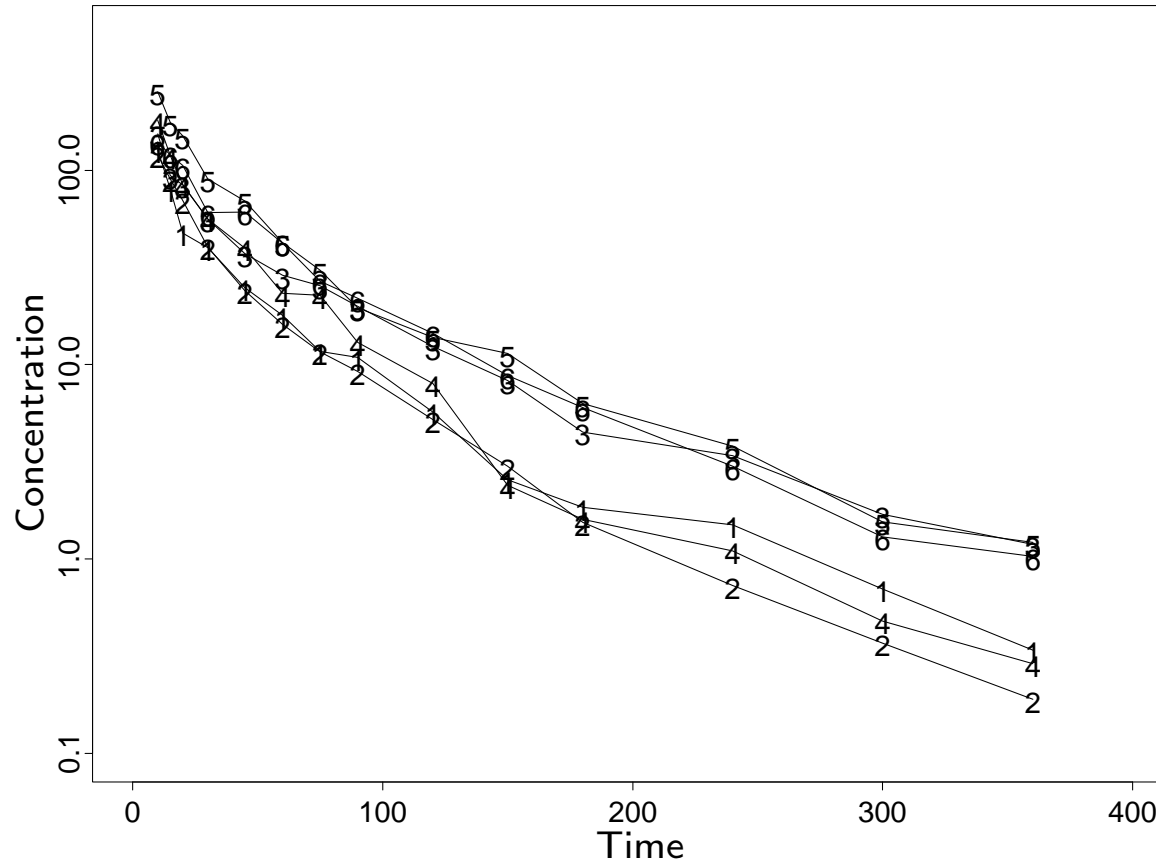
## Important:

- The PBPK model describes processes *within an single individual* giving rise to compartment-specific *concentrations for that individual*
- Parameters  $F$ ,  $V$ ,  $P$ ,  $V_{\max}$ ,  $K_m$  pertain to a *single individual*

## Fact of life:

- The *same* PK model may describe PK for any individual in the *population of interest*
- *However*, concentrations achieved during/following *identical exposure* may *differ substantially across individuals*
- $\Rightarrow$  Attributable to values of some or all of the *PK parameters* that *differ substantially across individuals*
- I.e., exhibit *variability* across *individuals in the population*

# Pharmacokinetic Models



*(See Section 2 of the paper for more on PK models)*

# Population PK Modeling and Analysis

---

**Perspective:** PBPK models in *risk assessment*

- The *scientific objectives* are relevant to *setting policy*
- ⇒ Interest focuses primarily on how PK mechanisms take place *overall in the population...*
- ...rather than in *any one specific individual*

**Thus:** Goal of *population PK modeling and analysis*

- Characterize formally how values of PK parameters *occur across the entire population*
- I.e., how the *are distributed* in the population ...
- ...and in particular how they *vary* in the population

# Population PK Modeling and Analysis

---

**Required:** A suitable *statistical model framework* in which *population variability* in PK parameters may be represented

- Forms the basis for *quantifying* the *distribution* of and *variability* in PK parameters across the population based on *data*

**Focus of analysis based on the statistical model framework:**

- *Estimate* quantities that describe the *distribution* of PK parameters in the population
- *Mean*, *variance* etc.
- “*Population parameters*”

**History:** Dating from the *1970s* for drugs

- Sheiner and Beal, UCSF

(*See Section 3 of the paper for more on population analysis*)

# Hierarchical Statistical Model

**Assume:** *Data* from an exposure study of  $N$  individuals,  $i = 1, \dots, N$

- On each individual,  $c$  *compartment-specific concentrations* measured *during/after exposure* at *times*  $t_{ij}$ ,  $j = 1, \dots, n_i$ , e.g.,  $C_{\text{ven}}$ ,  $C_{\text{exh}}$  ( $c = 2$ )
- Each individual  $i$ 's *exposure info* =  $E_i$
- $Y_{ijk}$  =  $k$ th concentration,  $k = 1, \dots, c$ , at time  $t_{ij}$  on individual  $i$
- $cn_i$  concentration measurements on  $i$  in a (*random*) *vector*

$$\mathbf{Y}_i = (Y_{i11}, \dots, Y_{in_i1}, \dots, Y_{i1c}, \dots, Y_{in_i c})'$$

- A *specific PBPK model* involving...
- *Physiological measurements*  $\phi_i$ , e.g., body weight, blood/air coefficient, lean body mass, etc, for individual  $i$
- *Unknown PK parameters*  $\theta_i$  for  $i$  (possibly *scaled*, *transformed*)

# Hierarchical Statistical Model

---

**Observed:**  $(Y_i, \phi_i)$  on each individual  $i = 1, \dots, N$  under *exposure conditions*  $E_i$  (*fixed by design*)

**Unobserved:** PK parameters  $\theta_i, i = 1, \dots, N$

**Assumption:** The  $(Y_i, \phi_i, \theta_i)$  arise (*statistically*) *independently*

- Individuals *sampled* from the *population* at *random*

**PBPK model:** *Solution* for  $k$ th concentration at time  $t$

$$f_k(t, E_i, \phi_i, \theta_i) \quad (\text{obtained } \textit{numerically})$$

- *Deterministic* function of time, PK, exposure

# Hierarchical Statistical Model

---

## Ingredients of the statistical model:

1. A *probability distribution* that describes how the *measured concentrations*  $Y_{ijk}$  take on their values *for individual  $i$* , *given* that individual  $i$  was exposed to  $E_i$  and has physiology  $\phi_i$  and PK  $\theta_i$
2. A *probability distribution* that describes how  $\theta_i$  are *distributed* (and hence *vary*) in the population

**Result:** A *two stage statistical model hierarchy* that *incorporates* the *deterministic PK model*

# Hierarchical Statistical Model

## Stage 1: Individual-level model

$$Y_{ijk} = f_k(t_{ij}, \mathbf{E}_i, \phi_i, \boldsymbol{\theta}_i) + e_{ijk} \quad \text{or}$$
$$\log(Y_{ijk}) = \log\{f_k(t_{ij}, \mathbf{E}_i, \phi_i, \boldsymbol{\theta}_i)\} + e_{ijk}$$

## Stage 2: Population model

$$\gamma_i = \log(\boldsymbol{\theta}_i) \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

- The *deterministic*, *mechanistic* PBPK model is *embedded* in the *individual-level statistical model*
- The “*deviations*”  $e_{ijk}$  represent the effects of phenomena *within an individual* that cause *measured concentrations* to *deviate* from the *deterministic* trajectory  $f_k(t, \mathbf{E}_i, \phi_i, \boldsymbol{\theta}_i)$
- $\boldsymbol{\mu}$ ,  $\boldsymbol{\Sigma}$  are *population parameters*

(See Section 4.1 of the paper for more on the basic model)

# Hierarchical Statistical Model

---

## Two sources of variability:

- *Intraindividual*
- *Interindividual*
- Describing *both of these* faithfully is *critical*
- $\Rightarrow$  Intra- and interindividual variability *combine* to produce *pattern of total variability*...
- ...so that *misrepresentation* of one leads to *misrepresentation* of the other!

**Thus:** Even though interest is in *the population* (*interindividual variability*), we *must* describe *intraindividual variability* correctly

# Hierarchical Statistical Model

---

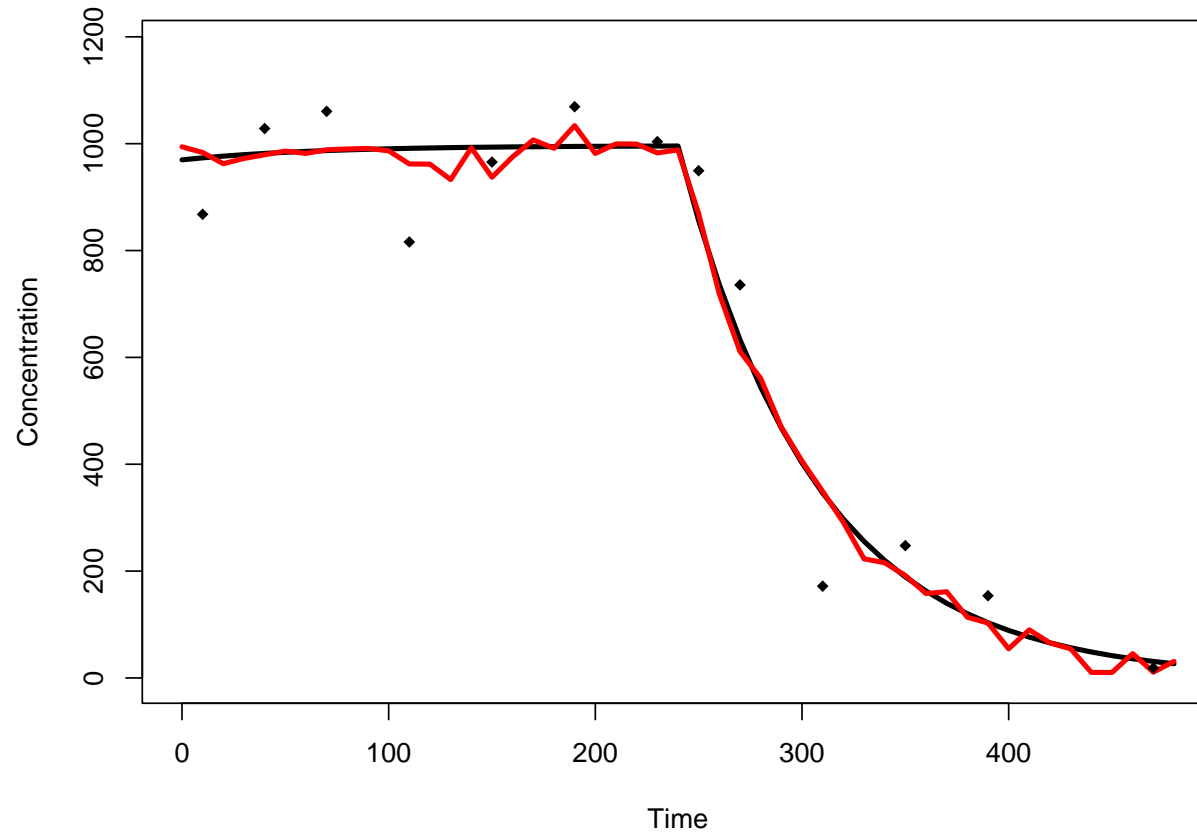
**Intraindividual variability:** Why do *measured concentrations* on individual  $i$  *deviate* from the *deterministic trajectory*?

- The PBPK model does not capture all within-individual physiological processes *perfectly* (e.g., variation in *breathing patterns*)
- Model assumptions aren't *perfectly met*  $\Rightarrow$  “*fluctuations*”
- *Measurement error* due to the *assay*

## Conceptual depiction:

- Solid black line = *deterministic PBPK model trajectory*
- Solid red line = *true concentration profile actually realized*
- $\blacklozenge$  = *measured values actually recorded* at intermittent times  $t_{ij}$

# Hierarchical Statistical Model



# Hierarchical Statistical Model

**Result:** *Interpretation* of  $f_k(t, \mathbf{E}_i, \phi_i, \boldsymbol{\theta}_i)$

- Average over *all possible realizations of true concentration profiles* and *measurement errors that could be committed* if *individual  $i$*  were observed under these conditions
- “*Inherent trajectory*” specifying the way  *$i$* ’s concentrations evolve over time “*on average*”

**Result:** *Two sources* of intraindividual variability

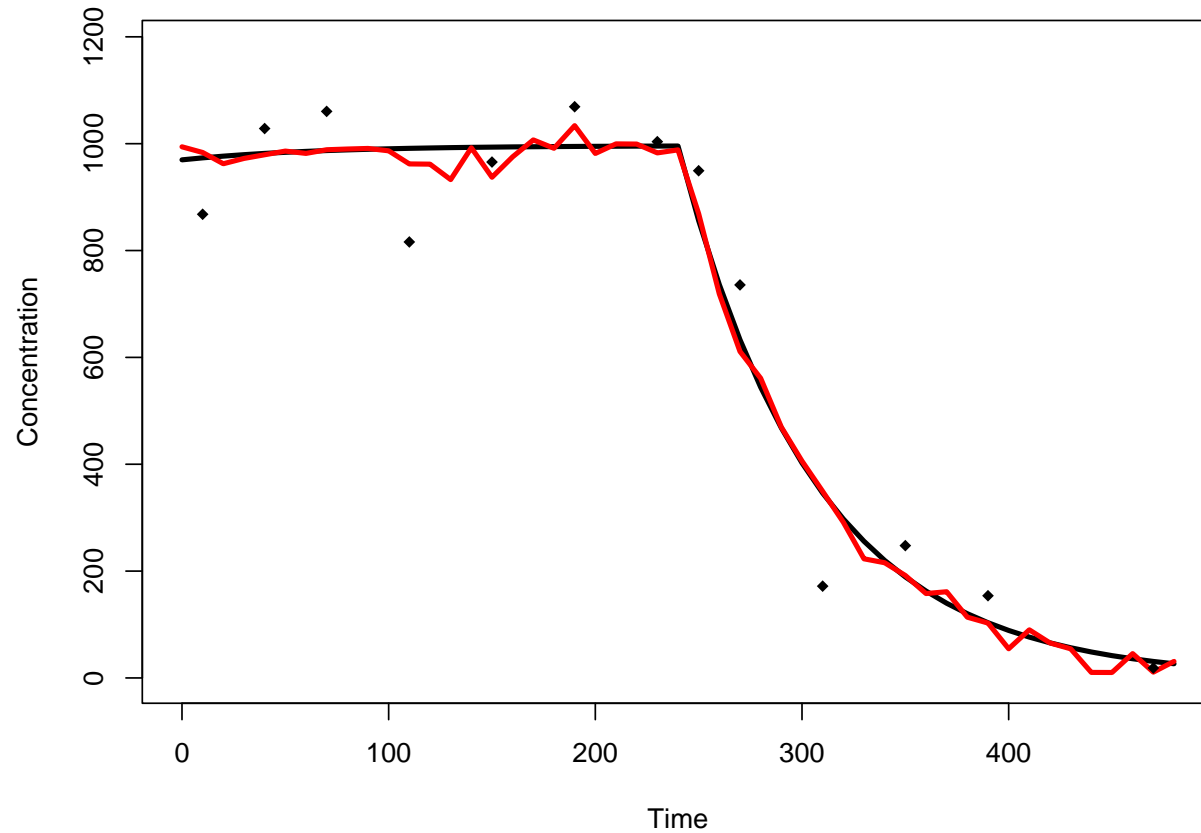
- “*Realization variability*” and “*measurement error variability*”

$$Y_{ijk} = f_k(t_{ij}, \mathbf{E}_i, \phi_i, \boldsymbol{\theta}_i) + e_{R,ijk} + e_{M,ijk} \quad \text{or}$$

$$\log(Y_{ijk}) = \log\{f_k(t_{ij}, \mathbf{E}_i, \phi_i, \boldsymbol{\theta}_i)\} + e_{R,ijk} + e_{M,ijk}$$

- Need *realistic assumptions* on  $e_{R,ijk}$ ,  $e_{M,ijk}$
- I.e., *Probability distributions* describing how they take on their values *separately and jointly*

# Hierarchical Statistical Model



# Hierarchical Statistical Model

**Considerations:** Holding  $\mathbf{E}_i, \phi_i, \theta_i$  fixed

- Measurement errors happen *haphazardly*  $\Rightarrow$  all  $e_{M,ijk}$  are *independent*  $\sim \mathcal{N}(0, \sigma_{M,k}^2)$
- Also  $e_{R,ijk} \sim \mathcal{N}(0, \sigma_{R,k}^2)$
- *However*,  $e_{R,ijk}$  *close together in time* may tend to “*deviate*” from  $f_k$  *in the same directions*  $\Rightarrow e_{M,ijk}$  are *correlated* for  $t_{ij}$  *close*
- ... and  $e_{R,ijk}$  for *different*  $k$  may also be *correlated*

**Usual assumptions:** *No correlations at all*

- *Result*: E.g., for approximate lognormality

$$\log(Y_{ijk}) | \mathbf{E}_i, \phi_i, \theta_i \sim \mathcal{N}[\log\{f_k(t_{ij}, \mathbf{E}_i, \phi_i, \theta_i)\}, \sigma_k^2], \quad \sigma_k^2 = \sigma_{R,k}^2 + \sigma_{M,k}^2$$

all *mutually independent*

- Is this *realistic*? *Consequences* if *not true*?

# Hierarchical Statistical Model

## Interindividual variability: *Population model*

$$\gamma_i = \log(\theta_i) \sim \mathcal{N}(\mu, \Sigma)$$

- $\mu = (\mu_1, \mu_2, \dots, \mu_p)'$  are *means in the population* of the  $p$  unknown  $\gamma_i$
- $\Sigma$  is the *covariance matrix*
- *Diagonal elements*  $(\Sigma_1^2, \Sigma_2^2, \dots, \Sigma_p^2)$  are the *variances in the population* of the  $p$   $\gamma_i$
- *Off-diagonal elements* are the *covariances in the population*, i.e., *correlations* – do individuals with high  $V_{\max,i}$  also tend to have high flow rate  $F_{\text{liv},i}$ ?

## Common assumption in PBPK population analyses: *No correlation*

- Is this *realistic*?

# Hierarchical Statistical Model

---

## Summary of the model:

- *Stage 1 individual-level model*: Probability distribution for how *measured concentrations*  $Y_{ijk}$  take on their values *for individual*  $i$ , *given* that individual  $i$  was exposed to  $E_i$  and has physiology  $\phi_i$  and PK  $\theta_i$  in terms of *intraindividual variability parameters*  $\sigma$
- *Stage 2 population model*: Probability distribution for how  $\theta_i$  ( $\gamma_i$ ) are *distributed* and *vary* in the population in terms of *population parameters*  $\mu$  and  $\Sigma$
- *Ideally* based on *scientifically justifiable assumptions* (not *convenience* or *simplicity*)

(*See Sections 4.2–4.7 of the paper for more on sources of variability and the model*)

# Hierarchical Statistical Model

---

**First goal:** *Estimate* the *population parameters*  $\mu$  and  $\Sigma$  from data from an exposure study

- Also *estimate*  $\sigma$  (or treat as *known* if dominated by *measurement error*)

**Important:** The statistical model is defined *independently of any estimation method*

- A representation of the processes by which *data are thought to arise*

# Variability and Uncertainty

---

**Variability:** Individuals and observations on them *vary* naturally

- *Interindividual variability* in particular is “*whatever it is*” – to make risk assessments, we need to *know about it*
- *Variability* in PK parameters  $\Rightarrow$  *variability* in *concentrations in target tissues*  $\Rightarrow$  *variability* in *adverse outcomes*

# Variability and Uncertainty

---

**Uncertainty:** *How well* can we learn about all this in the face of variability (and other issues)?

- One kind of uncertainty – is the *model correct*?
- *Given the model is approximately correct...*
- ...main kind of uncertainty, e.g., how “*reliable*” are *estimates* of population parameters?
- *Classical approach:* How much do results (estimates) change over *samples of data like the one we have*?
- Uncertainty measured by *sampling distribution* and its *covariance matrix*
- This uncertainty depends on *variability in the data* and  $N \Rightarrow$  to *reduce uncertainty*, need *larger  $N$*

(*See Section 4.8 of the paper for fuller discussion of variability and classical uncertainty*)

# Parameter Estimation

**Historically:** Estimation of  $\Omega = (\mu, \Sigma, \sigma)$  by *maximum likelihood*

- Based on *likelihood function* for  $\Omega$ ,  $L(\Omega)$ , given *observed data*
- $L(\Omega)$  describes the probability of “*seeing the data we saw*”  $\Rightarrow$  “*best estimates*” for  $\Omega$  maximize  $L(\Omega)$  (so maximize this probability)
- *Standard theory* gives approximate *sampling distribution* (for “*large N*”) that is *multivariate normal*; *sampling covariance matrix* quantifies (classical) uncertainty

# Parameter Estimation

**Practically speaking:** Maximizing  $L(\Omega)$  is *hard*

- When  $p$  is *large*  $\Rightarrow$  *high-dimensional integrals*, must be done *numerically*
- *Scant information* on some compartments  $\Rightarrow$  problems *identifying* components of  $\mu, \Sigma$
- *Numerical forward solutions* for  $f_k$  required

**Thus:** *Traditional* estimation methods try to *avoid* the integrals

- *Two-stage methods*: Get *individual estimates* of  $\theta_i$  or  $\gamma_i$  and use as “data” to estimate  $\mu, \Sigma$
- *First-order methods*: Approximate  $L(\Omega)$  to avoid the *intractable integrations*

(See Sections 5.2–5.7 of the paper for more on maximum likelihood and popular methods)

# Parameter Estimation

**Bayesian inference:** Another way of viewing *uncertainty*

- All model quantities ( $Y_i$ ,  $\gamma_i$ ,  $\mu$ ,  $\Sigma$ , etc) are *random vectors* with *probability distributions*
- *Probability distributions* for  $\mu$ ,  $\Sigma$  measure the *extent of uncertainty* about their values

**Ingredients:**

- The *hierarchical statistical model* describing how data arise
- A *prior distribution* for  $\Omega$  measuring our *uncertainty about their values in the absence of data*,  $p(\Omega) = p(\mu, \Sigma, \sigma)$
- A *posterior distribution* for  $\Omega$  updating our measure of *uncertainty about their values once we have seen data*
- The *mean* of the posterior distribution *estimates*  $\Omega$ ; its *covariance matrix* quantifies uncertainty

# Parameter Estimation

---

**Posterior distribution:** Can be obtained by *Bayes rule*

**Practically speaking:** Calculating the posterior is *hard*

- Involves *lots* of *high-dimensional integrals*
- *Same problems* as maximum likelihood

**Numerical method:** *Markov chain Monte Carlo* simulation

- Generate a large *sample of values* of  $\mu, \Sigma, \sigma, \gamma_i, i = 1, \dots, N$  from the *posterior*
- Use *means* as “*estimates*”
- *Important:* MCMC is a *computational technique*, has *nothing* to do with the statistical model

# Parameter Estimation

---

## Advantages:

- The *prior* is a *natural way* to incorporate information from *other sources*; e.g., literature, prior studies
- The *prior* can be used to *impose biologically-plausible constraints*
- Can help with *identification* where information is scant (will depend *solely* on prior)
- When info in data is “*good*” Bayesian and classical approaches give *comparable results*

## Criticisms:

- *Choice of prior* (*sensitivity*: different priors, different results?)
- *Computationally intensive* (no free lunch with any of this!)
- Some results *entirely predicated* on the *prior*
- “*Don't need large  $N$  to get answers*”

# Parameter Estimation

---

**Personal view:** For *population analysis*, using PBPK models, *Bayesian approach* is the most feasible

- Issues, limitations *must be understood* by users

(*See Sections 5.8–5.10 of the paper for more on Bayesian inference and the contrast with the classical view*)

# Concluding Remarks

---

## Topics for discussion:

- *Misspecification* – of the *PBPK model*? of the *statistical model*?  
How can we *tell*? What are the *consequences*? (*Sections 4.5, 6.1*)
- Why not use “*simpler pooled methods*?” – does not *respect the structure of the data* (*Section 5.1*)
- *Fixing parameter values* to circumvent *identifiability problems* – not a good idea! (*Section 5.1*)
- *Statistical model* for aggregated data? *Sample means and SDs*?  
From *serial sacrifice*? (*Sections 5.1, 6.1*)
- *Statistical model* for *combining data* from *multiple sources*?  
(*Section 6.1*)
- *Model extensions*; e.g., *interoccasion variability* (*Section 6.1*)

# Concluding Remarks

---

**Implications:** Estimation of *population PK distribution* is interesting, but *not an end unto itself*

- *Risk estimates* based on results
- *Variability* in PK parameters  $\Rightarrow$  *variability* in *concentrations in target tissues*  $\Rightarrow$  *variability* in *adverse outcomes*
- *Uncertainty* about PK parameters  $\Rightarrow$  *uncertainty* about *concentrations in target tissues*  $\Rightarrow$  *uncertainty* about *adverse outcomes*

**Critical:** The *hierarchical statistical model* provides a *formal basis* for characterizing this variability and uncertainty

(*See Section 6.2 of the paper*)