The role of methodology and statistics in research integrity

Geert Molenberghs

geert.molenberghs@uhasselt.be & geert.molenberghs@kuleuven.be

Interuniversity Institute for Biostatistics and statistical Bioinformatics (I-BioStat)

UHasselt & KU Leuven, Belgium

www.ibiostat.be

KVAB, October 18, 2017

UHASSELT

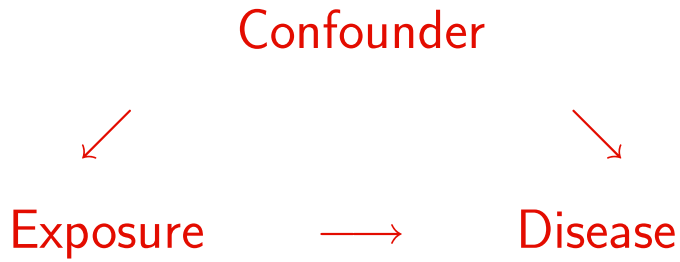
I-BioStat

KU LEUVEN

Interuniversity Institute for Biostatistics
and statistical Bioinformatics

Theme 1
Research and Research Conduct

1.1 Observational Studies: Environment and Health



- **Environment & health:** Waste incineration (dioxin) and congenital malformation
- **Smoking & lung cancer:** tobacco industry versus the states of the US
- **Cadmium in the north of Limburg**

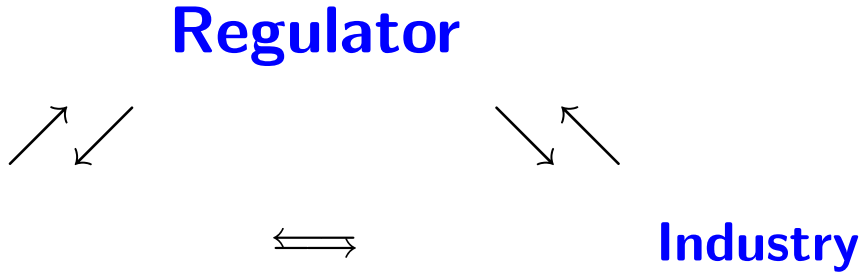
1.2 Further Problems with Experiments: Psychology, Sociology, Economy, Medicine, Exact Science...

! Not done: invention of studies/study results

? Not done: removal of study subjects that do not fit well with the rest

- $243 \times 3 = 729$
- SOLAS
- 67,138 BEF

1.3 An Experimental Setting: Randomized Clinical Trials



- Phases of clinical research
- Randomization
- Blindness
- Various committees:
  - **Institutional Review Board**
  - **Data Monitoring Committee**
- Informed consent

Theme 2
Case Study: The Toenail Data

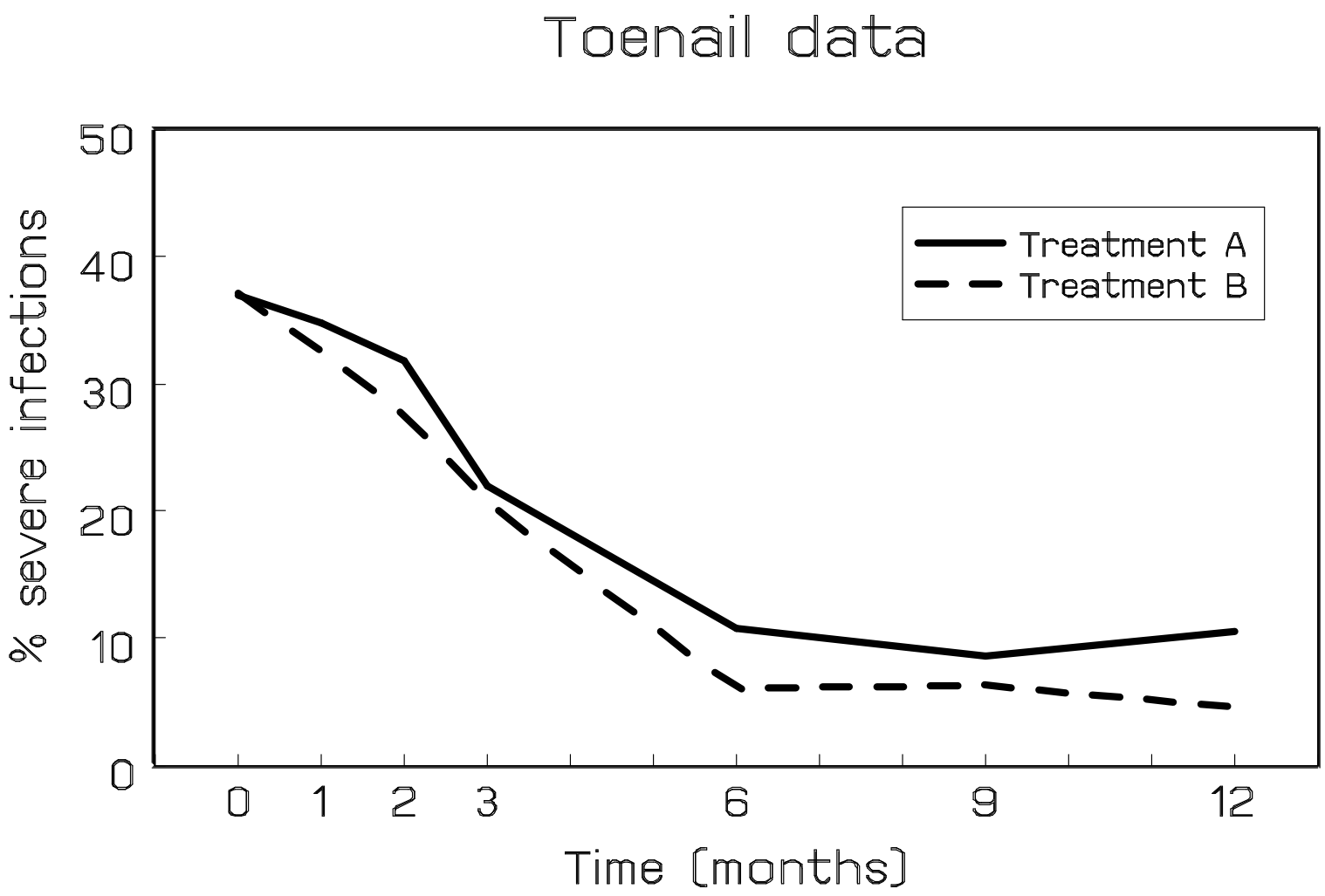
- **T**oenail **D**ermatophyte **O**nychomycosis: Common toenail infection, difficult to treat, affecting more than 2% of population.
- Classical treatments with antifungal compounds need to be administered until the whole nail has grown out healthy.
- New compounds have been developed which reduce treatment to 3 months
- Randomized, double-blind, parallel group, multicenter study for the comparison of two such new compounds (A and B) for oral treatment.

- Research question:

Severity relative to treatment of TDO ?

- 2×189 patients randomized
- 48 weeks of total follow up (12 months)
- 12 weeks of treatment (3 months)
- measurements at months 0, 1, 2, 3, 6, 9, 12.

- Frequencies at each visit (both treatments):



2.1 Application to the Toenail Data

- Consider the model:

$$
\begin{aligned}
Y_{ij} &\sim \text{Bernoulli}(\mu_{ij}) \\
\log\left(\frac{\mu_{ij}}{1-\mu_{ij}}\right) &= \beta_0 + \beta_1 T_i + \beta_2 t_{ij} + \boxed{\beta_3 T_i t_{ij}} \\
\text{Corr}(Y_{ij}, Y_{ij'}) &= \alpha \qquad \text{(working correlation)}
\end{aligned}
$$

- Y_{ij}: severe infection (yes/no) at occasion j for patient i
- t_{ij}: measurement time for occasion j
- T_i: treatment group

2.1.1 Inference on Key Parameter: β_3. Story 1.

Model	Estimate (s.e.)	p-value
Initial model	-0.0783 (0.0394)	0.0469
Model-based (naive)	-0.0886 (0.0362)	0.0143
Empirically corrected (robust)	-0.0886 (0.0571)	0.1208

"The initial model is the most efficient estimator, because it assumes that each data point provides an independent piece of information. Based on this model, the treatment effect is marginally significant."

2.1.2 Inference on Key Parameter: β_3. Story 2.

Model	Estimate (s.e.)	p-value
Initial model	-0.0783 (0.0394)	0.0469
Model-based (naive)	-0.0886 (0.0362)	0.0143
Empirically corrected (robust)	-0.0886 (0.0571)	0.1208

"The model-based estimator assumes that the various pairs of measurements per patient exhibit a common correlation. This is estimated to be $\widehat{\alpha} = 0.42$*, considered to be a plausible value. Therefore, inferences are based on the model-based estimator; this leads to a significant effect of treatment, with* $p = 0.0143$*."*

2.1.3 Inference on Key Parameter: β_3. Story 3.

Model	Estimate (s.e.)	p-value
Initial model	-0.0783 (0.0394)	0.0469
Model-based (naive)	-0.0886 (0.0362)	0.0143
Empirically corrected (robust)	-0.0886 (0.0571)	0.1208

"The empirically-corrected estimator assumes that the various pairs of measurements per patient exhibit a common correlation, but that, at the same time, this correlation assumption may be incorrect. In other words, it protects against misspecification. Inferences are based on this estimator. We conclude that there is no significant effect of treatment, with $p = 0.1208$*."*

2.1.4 Inference on Key Parameter: β_3. Story 4.

Model	Working corr. α	Estimate (s.e.)	p-value
Initial model		-0.078 (0.039)	0.0469
Model-based (naive)	exchangeable	-0.089 (0.036)	0.0143
Emp. corr. (robust)	independence	-0.078 (0.055)	0.1515
Emp. corr. (robust)	exchangeable	-0.089 (0.057)	0.1208
Emp. corr. (robust)	unstructured	-0.114 (0.052)	0.0275

"The empirically-corrected estimator assumes that the various pairs of measurements per patient exhibit a certain structure, but that, at the same time, this correlation assumption may be incorrect. The working correlation that is closest to the true structure is generally most efficient. Inferences are based on this estimator, with unstructured working correlation. We conclude that there is a significant effect of treatment, with $p = 0.0275$*."*

2.2 The Generalized Estimating Equations Case: Discussion

- Nice method to efficiently and correctly analyze non-Gaussian longitudinal data

- But: there are pitfalls

  ▷ Initial $\longleftrightarrow$ Model-based $\longleftrightarrow$ Empirically corrected

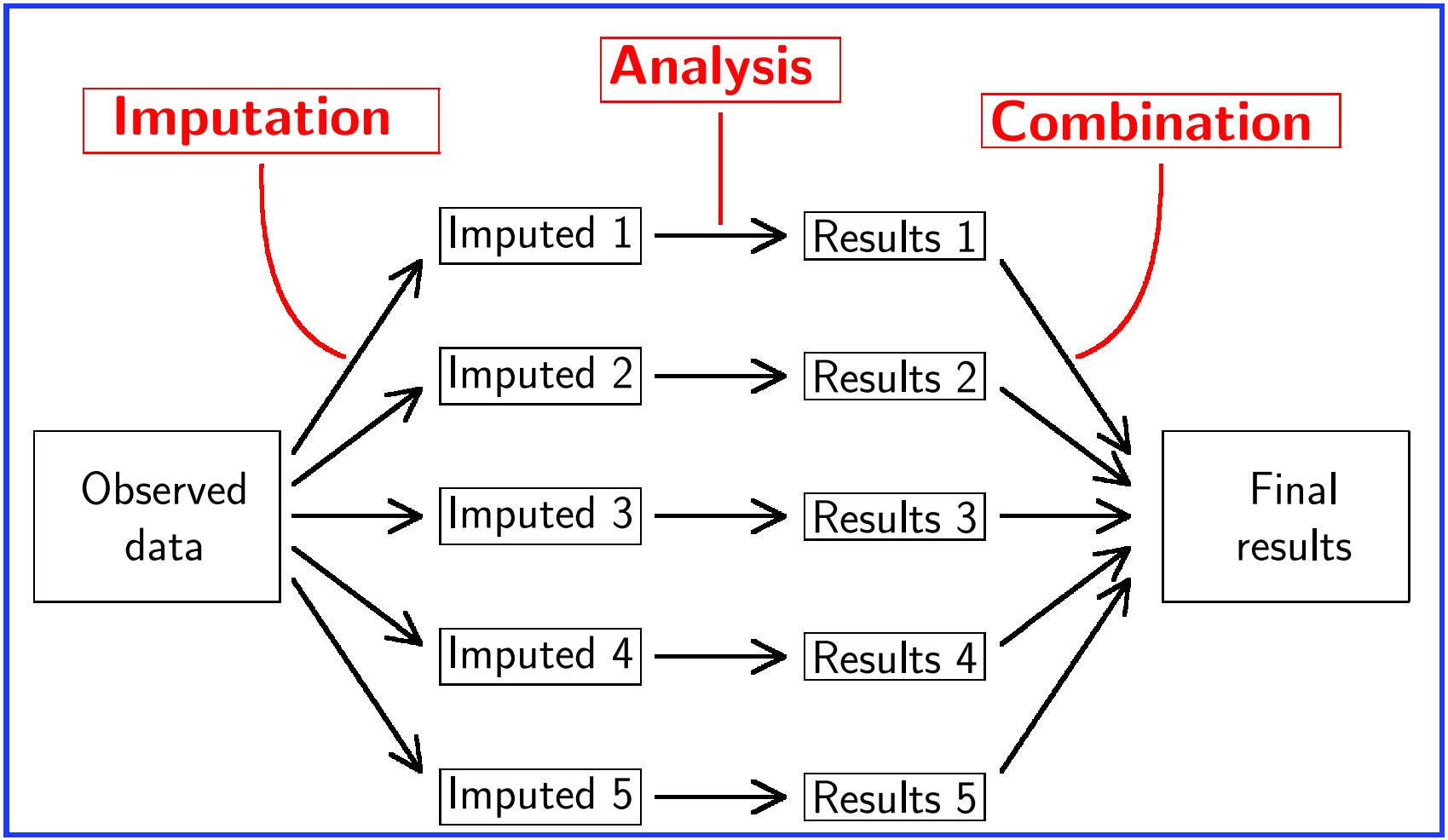
  ▷ Various working correlation structures possible: **how** and **when** to choose?

- **Know the method! Know the pitfalls! Stand firm on principles!**

Theme 3
Multiple Imputation

- Incompletely observed repeated measures
- A procedure gaining a lot of clout,...
- Three steps:
  1. The missing values are **sampled** M times $\Longrightarrow$ M complete data sets
  2. The M complete data sets are analyzed by using standard procedures
  3. The results from the M analyses are combined into a single inference
- Rubin (1987), Rubin and Schenker (1986), Little and Rubin (1987)

- Multiple imputation ($M = 5$ imputations):



3.1 Multiple Imputation: Ticket to Fraud?

- Code for imputations:

```
proc mi data=armd13 seed=486048 out=armd13a simple nimpute=10 round=0.1;
   var lesion diff4 diff12 diff24 diff52;
   by treat;
   run;
```

- ?
- **seed=486048**
- !

Theme 4
Scientific Integrity, Ethical Conduct, or Not Always?

- **Methodological contributions:**
  - ▷ Plagiarism detection methodology
  - ▷ General fraud detection methodology
- **Checking:**
  - ▷ Reproducible research — software / data
- **Critical reflection:**
  - ▷ Examination of measures and metrics:
  - ▷ $\Delta p \cdot \Delta r \geq \overline{h}$
  - ▷ Bye impact factor!