

# Phases of methodological research – and the role of simulations

Georg Heinze,

Anne-Laure Boulesteix, Michael Kammer, Tim Morris, Ian White

for the Simulation Panel of the STRATOS Initiative

# ,Rome wasn't built in a day'



Colosseum 72-80



Pantheon 114-126



St Peter 1506-1626

Who originally said Rome wasn't built in a day? ^

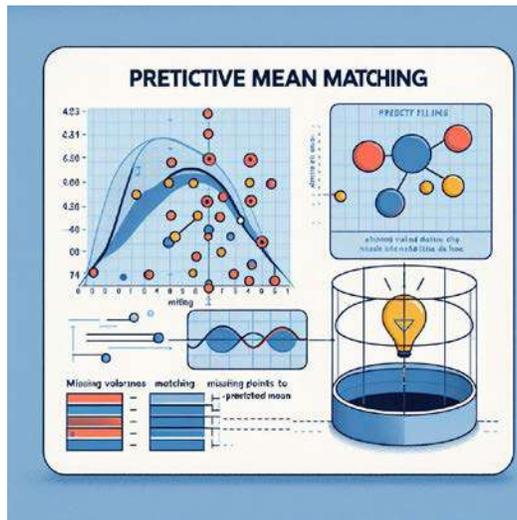
John Heywood's

English playwright, [John Heywood's](#) saying that “Rome wasn't built in a day, but they were laying bricks every hour”, is a reminder of the fact that it requires time and patience to create something big and great. (Google)

# Predictive mean matching: 1988-2014

Little (1988): Proposed predictive mean matching (PMM).  
Linked to multiple imputation.  
No proof.

Heitjan and Little (1991) – applied PMM and performed little simulation study. Promising though not ideal performance.



Schenker and Taylor (1996) – reasonably broad simulation study with 3 variants of PMM + other methods

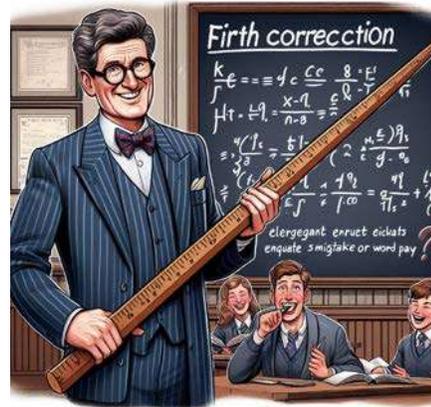
Morris et al (2014):  
Review of PMM literature  
Simulation study to clarify when PMM performs poorly and when it outperforms other methods

# The Firth correction: 1993-2018

Firth (1993): Bias reduction of maximum likelihood estimates

Heinze and Schemper (2002): A solution to the problem of separation in logistic regression

van Smeden et al (2016): No rationale for 1 variable per 10 events criterion for binary logistic regression analysis



Puhr et al (2017): Firth's logistic regression with rare events: accurate effect estimates and predictions?

Mansournia et al (2018): Separation in logistic regression: Causes, consequences and control

# Drug development needs time



<b><u>Duration</u></b>	3-6 years since the drug discovery	Months	Months-Years	Years	Ongoing following FDA approval
<b><u>Sample size</u></b>	Not specific	Small	Large	Larger	Impacting larger beyond
<b><u>Population</u></b>	In vitro and In vivo animals	Healthy population / may be with targeted disease e.g. cancer, T.B. etc.	Population with target disease	Diverse population with target disease	Diverse population with target disease & new age groups, gender
<b><u>Types of studies</u></b>	Not specific	Unblinded & Uncontrolled	May be Placebo (inactive substance) & Active, Controlled	Randomized & controlled	Expanded safety comparison
<b><u>Factors to be identified</u></b>	Mechanism of action, Efficacy (ability of drug to act against pathogen/disease), Safety (concerns the medical risk to the patient)	Pharmacodynamics (side effect / desire effect, mechanism of action of drug) Pharmacokinetics (absorption, distribution, metabolism & elimination of drug), Tolerated dose	Drug-Drug & Drug-Disease interaction, Efficacy at various doses, Patient safety	Dosage intervals, Risk-benefit information, Efficacy and safety for subgroups	Epidemiological data, Efficacy and safety within large diverse populations, Pharmacoeconomics (comparison of value of one pharmaceutical drug or drug therapy to another)

\*Tolerability: Represents the degree to which overt adverse effects can be tolerated by the patient.

\*\*Effectiveness: The extent to which a drug achieves its intended effects.

**References:**

1. <http://www.nlm.nih.gov/services/ctphases.html>
2. <http://www.fda.gov/drugs/resourcesforyou/consumers/ucm143534.htm>
3. [http://www.innovation.org/drug\\_discovery/objects/pdf/RD\\_Brochure.pdf](http://www.innovation.org/drug_discovery/objects/pdf/RD_Brochure.pdf)

# Learning from drug development

## Phases of research as a framework for building evidence

Drug development



Biostatistical methods



# Methodological research: Phase I

- Aim:
  - Introduce new idea to solve a problem
  - Demonstrate its validity by investigating its properties,
  - Show potential to improve on existing solutions or to be the only solution
- Elements of a study:
  - Mathematical derivations and proofs
  - Simple example data analyses
- After that phase we know:
  - Whether method is valid or invalid from a theoretical point of view

# Methodological research: Phase II

- Aim:
  - Demonstrate use of method with real data
  - Introduce refinements and extensions
  - Considering a limited range of possible applications
- Elements of a study:
  - Simulations including limited comparisons with other methods
  - Simple example data analyses
- After that phase we know:
  - Whether method can be used with caution  
or should not be used in certain applied settings

# Methodological research: Phase III

- Aim:
  - Comparing a relatively new method with competitors
  - Demonstrating its use in practice
  - Considering a wide range of applications
- Elements of a study:
  - Refinements of method to broaden applicability
  - Simulations with a wide range of scenarios and different outcome types, set up as neutral comparison studies
  - Realistic comparative example data analyses
- After that phase we know:
  - In which settings (among many) a method can be safely used
  - In which settings it outperforms other methods

# Methodological research: Phase IV

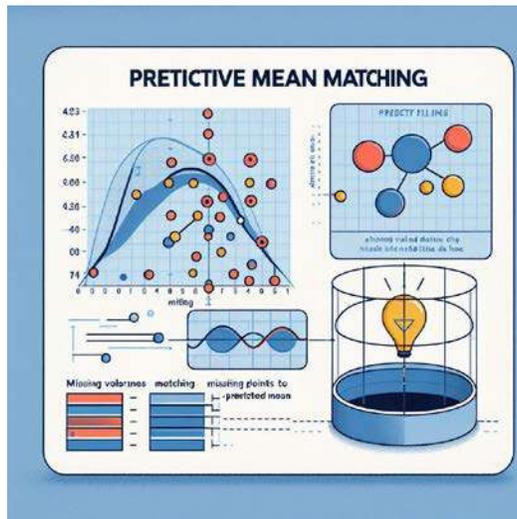
- Aim:
  - Summarizing the evidence about a method, also in comparison with competing methods
  - Uncovering previously unknown behaviour with complex data
  - Considering an extended range of possible and actual applications
- Elements of a study:
  - Review of existing evidence about a method
  - Simulations with extended range of scenarios
  - Complex comparative example data analyses
- After that phase we know:
  - When a method is preferred and when it is not
  - What diagnostics are available
  - Which pitfalls may occur with its application

# The role of simulations and synthetic data

# Predictive mean matching: 1988-2014

Little (1988): Proposed predictive mean matching (PMM).  
Linked to multiple imputation.  
No proof.

Heitjan and Little (1991) – applied PMM and performed **little simulation** study. Promising though not ideal performance.



Schenker and Taylor (1996) – reasonably **broad simulation** study with 3 variants of PMM + other methods

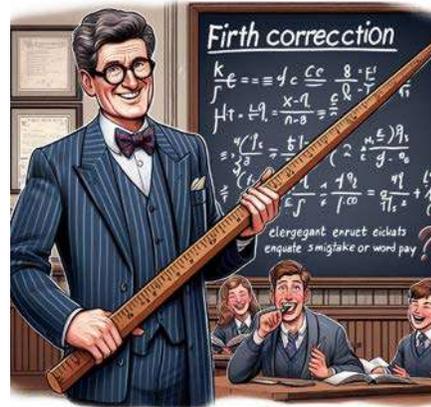
Morris et al (2014):  
Review of PMM literature  
**Simulation study to clarify** when PMM performs poorly and when it outperforms other methods

# The Firth correction: 1993-2018

Firth (1993): Bias reduction of maximum likelihood estimates

Heinze and Schemper (2002): A solution to the problem of separation in logistic regression (simple simulations)

van Smeden et al (2016): No rationale for 1 variable per 10 events criterion for binary logistic regression analysis (broad simulation study)



Puhr et al (2017): Firth's logistic regression with rare events: accurate effect estimates and predictions? (simulations)

Mansournia et al (2018): Separation in logistic regression: Causes, consequences and control

# The role of simulations and synthetic data

- **Phase I:** single ,toy‘ examples – **synthetic** data
  - Used to demonstrate how the method works
- **Phase II:** limited range of **simulation** scenarios
  - To compare the method with (selected) others
  - Still ,inventor-biased‘
- **Phase III:** broad range of **simulation** scenarios
  - ,Neutral‘, wide range of applications in mind
- **Phase IV:** establishing trustworthiness or finding breakdown scenarios
  - May be very wide **simulation** studies
  - Could be focused on single but likely ,difficult scenarios‘ (**synthetic** data sets?)
  - May focus on diagnostics for safe application: when is the method preferred?

Development

Evaluation, multiple methods

Evaluation or review:  
one special scenario – many methods  
one method – many scenarios

# *Don't expect anyone to be able to build Rome in one day!*

- Authors of methods research should **clearly disclose the phase** they're contributing to!
- For Phase I: do not ask authors to prove that their new method works in all hypothetical scenario; allow ,high-risk' methods
- For Phase II: reduce the risk: comparison included? Data example?
- For Phase III-IV: specifically check neutrality and broadness of comparisons. Realistic data example?
- For Phase IV: is it clear when the method is to preferred over others and when not?
- *R*

# Also Vienna wasn't built in a day



1359-1439  
(tower)



1696-1705, 1742-1749



1964-1994



1896-1897

2020-2026\*



## Phases of methodological research in biostatistics—Building the evidence base for new methods



Georg Heinze<sup>1</sup> | Anne-Laure Boulesteix<sup>2</sup> | Michael Kammer<sup>1,3</sup> | Tim P. Morris<sup>4</sup> |  
Ian R. White<sup>4</sup> | on behalf of the Simulation Panel of the STRATOS initiative

HOME

ABOUT 

CONTRIBUTE 

BROWSE 

VIRTUAL ISSUES

### Special Collection: “Neutral Comparison Studies in Methodological Research”

Virtual Issues | First published: 14 December 2023 | Last updated: 19 February 2024

Biometricians are frequently faced with a multitude of methods they might use for the analysis and/or design of studies. Choosing an appropriate method is a challenge, and neutral comparison studies are an essential step towards providing practical guidance. This Special Collection contains both papers defining, developing, discussing or illustrating concepts related to the design and interpretation of neutral comparison studies, and reports of neutral comparison studies of methods that address specific biostatistical problems.

**Guest editors:** Anne-Laure Boulesteix, Mark Baillie, Dominic Edelmann, Leonhard Held, Tim Morris, Willi Sauerbrei

# The phases of methodological research

**TABLE 1** A brief description of the proposed scheme of phases of methodological research

Phase	Scope: A study in that phase will typically aim at ...	Elements: Typically, a study in that phase will consist of...	Outcome: after that phase, we know...
I	... introducing a new idea, demonstrating its validity by investigation of (asymptotic or finite-sample) properties, showing potential to improve on existing methods or to be the only solution.	... mathematical derivations and proofs, very simple example data analyses.	... whether a method is valid or invalid from a theoretical point of view.
II	... demonstrating the use of the method with real data, probably introducing refinements and extensions; it will consider only a limited range of possible applications.	... simulations including limited comparisons with other methods, simple example data analyses.	... whether a method can be used with caution or should not be used in certain applied settings.
III	... comparing a relatively new method with competitors and demonstrating its use in practice; it will consider a wide range of applications.	... simulations with wide range of scenarios and different outcome types (ideally set up as neutral comparison studies), realistic comparative example data analyses.	... in which settings (among many) a method can be safely used and in which it outperforms competing methods.
IV	... summarizing the evidence about a method, also in comparison with competing methods; uncovering previously unknown behavior of the method in complex data analyses; considering an extended range of possible and actual applications.	... a review of the existing evidence about a method, simulations with extended range of scenarios, complex comparative example data analyses.	... when a method is and when it is not the preferred method; what diagnostics are available and which pitfalls may occur with its application.

Heinze et al,  
Biometrical Journal 2023