How biostatistical methods mature: understanding the four phases of methodological research

Georg Heinze

Institute of Clinical Biometrics

Credits

STRATOS INITIATIVE

DOI: 10.1002/bimj.202200222

Biometrical Journal

RESEARCH ARTICLE

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

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

Anne-Laure Boulesteix

Michael Kammer

Tim Morris

Ian White

Motivation

- Many new biostatistical methods are being proposed
- Few of them are used in practice
- If specific methods are used, can we trust them?
  - Which methods can we safely use in data analyses?
  - Which methods still lack appropriate validation or have hidden pitfalls?

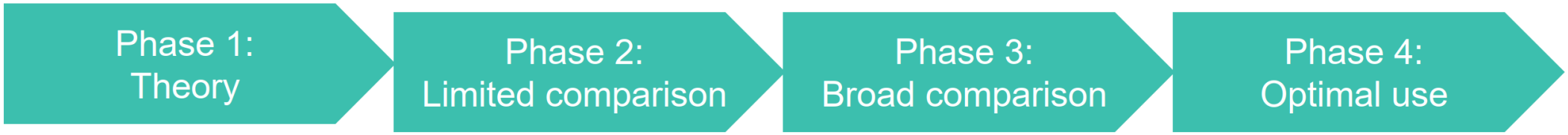
- → We introduced a framework that looks at biostatistical method development similarly as drug development in clinical medicine

Learning from drug development

Phases of research as a framework for building evidence

Drug development

Biometrical methods



Heinze, Boulesteix et al,
Biometrical Journal 2024

The four phases

Phase	Scope	Typical Activities	Outcome
I	**Introduction** of a new method	Theoretical development, proofs, asymptotics, basic illustrations	Demonstrates theoretical validity
II	**Initial application and evaluation**	Limited simulations, real-life applications that are not too complex (with „cleaned data“, ...). Inventor usually involved	Demonstrates usefulness, but still with caution, restricted to the specific investigated settings
III	**Broader evaluation and comparison** of still relatively new method (compared to other probably established or new methods)	Neutral comparison studies (inventor bias avoided or transparently disclosed), extensive simulations, diverse real-world examples	Comparative performance, strengths, limitations
IV	**Evidence synthesis and increased understanding** about a method that has been in use for some time (‚post-marketing‘)	Reviews, complex applications, wide simulations in new, previously unconsidered settings, identification of possible pitfalls, development of diagnostics for a method	Clarifies when the method is preferable over others, or comparable to others, or when it should not be used and why

What each phase adds

- Phase I:
  lays theoretical basis, toy examples
- Phase II:
  adds real example, limited comparisons (comparator methods/DGMs),
  still inventor-biased
- Phase III:
  comparisons become broad and neutral
- Phase IV:
  after many years in use: extended comparisons, evidence synthesis, detection and description of possible pitfalls in application,
  when is the method (not) preferred?

Can we easily classify papers/studies/implementations into phases?

- Some experience:
- Early phases focus on method development,
  later phases on evaluations/comparisons for specific applications
- Phase I and II are separated in our framework, but are often combined in practice
- Most studies *contribute* to a phase but *do not conclude* it
- Distinguish phase of development of *method* from development of *software*
- Feedback loop: detection of pitfalls in Phase IV may lead to proposal of a correction/modification, which itself starts again its lifecycle

Assessing papers for their phases

- We assessed some papers manually and with the use of AI:

- Many papers published in biostatistical journals can be assigned to Phase II
- Few studies are pure Phase I (probably found in more theoretical journals)
- Few studies Phase III, even fewer Phase IV

- → The discipline is heavily weighted toward early phases of methodological development

Consequences of the dominance of new methods

- Lack of full understanding about new methods
- Over-confidence when using them
  - particularly when *their inventors* are using them
- Practictioners would need critical, evidence based guidance for safe selection and application of proper methods
  - But this evidence base is often lacking
- Uncritical application of new methods may have lead to widespread errors in interpretation (of medical studies)
- Also the opposite may be true: methods may have been said to perform poorly, but are actually robust choices

Recommendations for the biostatistical community (1)

- Researchers:
  - Prioritize well-planned Phase III and IV studies!
    - Don‘t take it easy, these studies are tough!
    - Publish your simulation protocols, let the community participate
  - Emphasize the strengths, but don‘t hide pitfalls and shortcomings of methods!
  - Embrace ‚adversarial collaborations‘ to ensure rigorous and unbiased comparisons!

Recommendations for the biostatistical community (2)

- Journal editors and reviewers:
  - Still, proposals of new methods are considered more ‚innovative' than comparisons of existing methods (‚*yet another simulation paper*')
  - But: comprehensive evaluation of existing methods is very valuable for the health of the discipline!
  - Actively solicit and prioritize Phase III and IV studies, and tutorials and reviews!
  - Let authors declare which phase (they think) their research contributes to!
  - Let reviewers assess which phase (they think) a paper contributes to!

Recommendations for the biostatistical community (3)

- Practitioners:
  - Use our framework as a lens to critically evaluate new methods before adoption!
  - Foster a culture of methodological quality from the ground up!
    - Provide the reasons why in an analysis you preferred a specific method over others!
      - Is there a solid evidence base for your decision?
    - Register analysis protocols
    - Enable reproducibility of analyses (share: code, data, IDA results, history of SAP, …)
    - Follow reporting guidelines

Three fine examples for methods development and evaluation

- Michal Abrahamowicz: Weighted cumulative exposure modeling
- Willi Sauerbrei: MFPI
- Ewout Steyerberg: Net benefit and Net reclassification index

Putting the examples into the phases context

- When and how were negative aspects of the method discovered and published?
- How was neutrality seeked for in phase 3 and phase 4?
- When and how was the method first used in applications?
  - By the inventors or by other researchers?
  - At which phase?