Methods and Guidelines for Integrity of Multivariable Analysis of Real-World (Observational) Data

Poster · October 2014

2 authors:

Aaron Kamauu
Ikaika Health

Andrew Wilson
PAREXEL International Corp.

Methods and Guidelines for Integrity of Multivariable Analysis of Real-World (Observational) Data

DG Levy,[1] D Norris,[2] A Kamauu,[3] A Wilson[3]

[1]Evidence Science & Innovation—US Medical Affairs, Genentech, Inc.; [2]David Norris Consulting, LLC; [3]Anollinx, LLC

BACKGROUND

A consequence of the proliferation of “real-world” data (i.e., non-randomized studies; e.g., electronic health records, claims data, public-use data, ACOs, etc.) and the urgency to use them to inform clinical decision making is that the prevalence of observational data analyses (ODA) is increasing dramatically. The methods, standards and guidelines for multivariable ODA currently borrow substantially from those for randomized trials and do not specifically address idiosyncrasies of ODA important for reproducible and scientifically meaningful results. Concern about the credibility of ODA for scientific evidence is widespread. Updated approaches to addressing ODA credibility are needed.

OBJECTIVES

- To critically examine standard practice of ODA and survey the manifold opportunities for misdirection of reproducible valid inference.
- Guidelines for conducting ODA and corresponding standards for evaluating ODA are proposed that integrate corrective propositions.

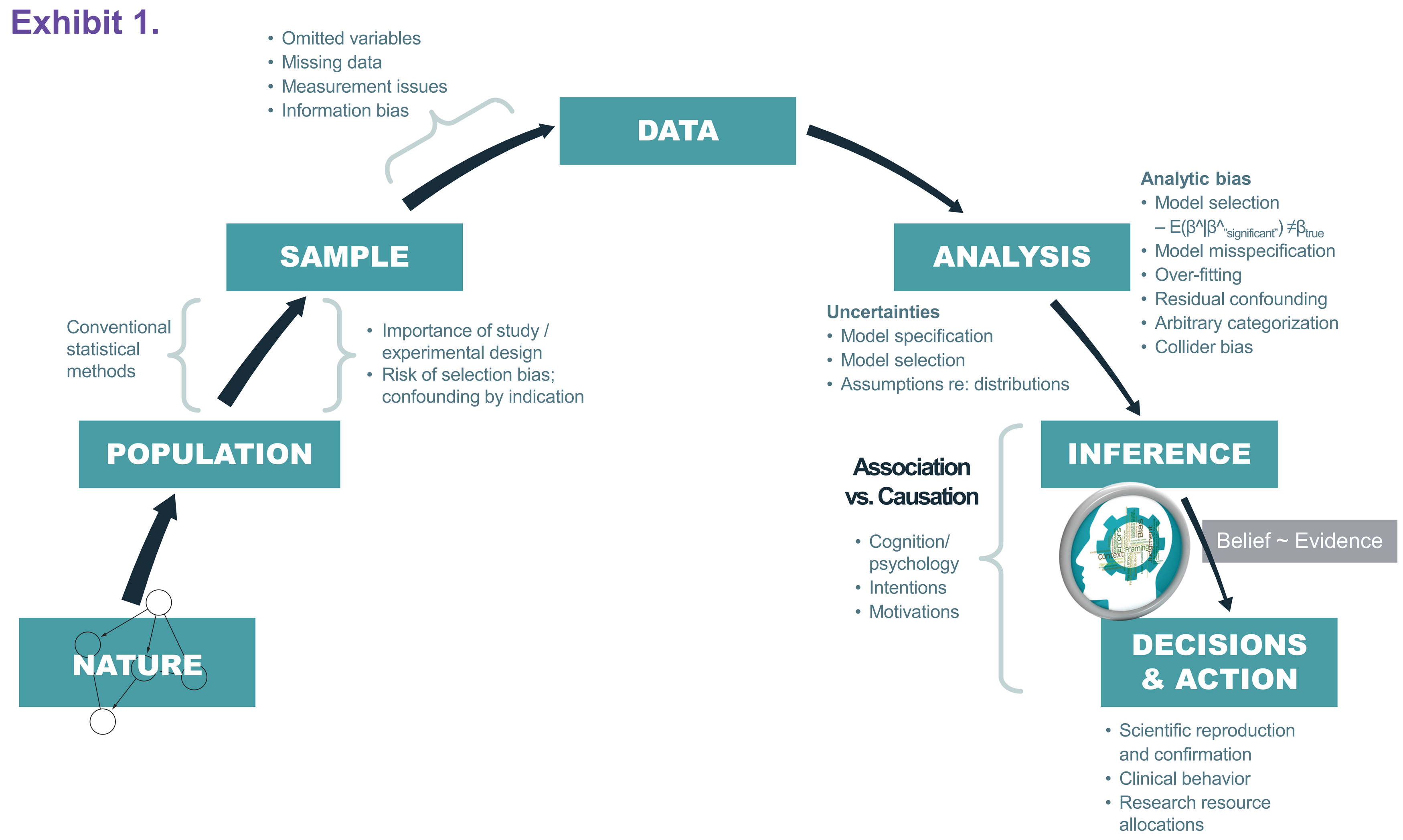
METHODS

The entire process generating scientific evidence via ODA is deconstructed and analyzed for prevalent sources of potential distortion of information and inference, and recommendations for remediation at critical points are made. Criteria that support the self-correcting mechanism in science are emphasized. Major points of focus include (a) explicit specification of the underlying causal and modeling assumptions; (b) uncertainties due to model specification, over-fitting, predictive optimism; (c) formally incorporating reproducibility and expectation for reproduction of research into ODA; and (d) expression and evaluation of results in terms of reduction of uncertainty and altering prior belief.

RESULTS

- The process of generating scientific evidence via ODA is characterized (Exhibit 1) in its entirety, and sources of information loss or distortion are identified.
- An analogy is made between sources of information loss or distortion in the ODA process and noise in a communication channel (Exhibit 2).
- Components of uncertainty that accrue in conventional ODA are suggested (Exhibit 3).
- Criteria that support the self-correcting mechanism in science are emphasized (Exhibit 4).

Exhibit 1.



“What we observe is not nature itself, but nature exposed to our method of questioning.”

- Werner Heisenberg

INFORMATION CHANNEL: Components

Nature: The complex and nexus of causes that produce the phenomena of our world that are available for empirical study. The underlying causal structure of nature is often abstruse or inscrutable.

Population: All of the objects (existing, nonextant and/or possible) in the category of interest for study. The population is the realization of causal process in nature. The population is also the primary object of study and inference.

Sample: The subset of the population available for study and observed.

Data: The actual observations made and recorded on the sample. Not all observations/variables of all possible variables from the sample are collected. Measurements are made imperfectly and recorded with errors. The particular instance of the data (out of many possible instances) inform the likelihood on which the analysis is predicated.

Analysis: The mathematical procedures that account for both the structure and randomness of the data. Typically a model is used or is at least implicit. All analyses require assumptions (both strong and weak).

Inference and Belief: The conclusions drawn from the analysis of the data (and in combination with any external information), including whether any associations observed are causal and are likely to be reproducible effects in independent data. Belief depends on the strength of the findings and the research process, coherence with existing knowledge, and numerous cognitive and psychological factors, including biases, intentions and motivations.

Decisions and Actions: The consequences, if any, of the research activities. The impact of the research will depend in part on the strength of the belief resulting from the inference and the relevance for problems faced by others. Consequences include clinical behavior and medical decision making; and scientific behavior, including confirmatory reproduction of research and motivation of additional research.

Exhibit 2.
Schematic diagram of a general communication system

The sources of information loss or distortion are analogous to sources of noise[1] in a communication channel.

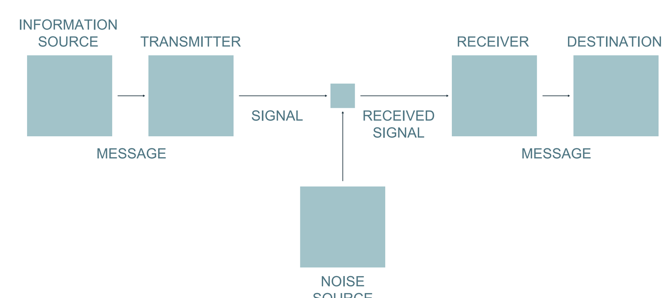
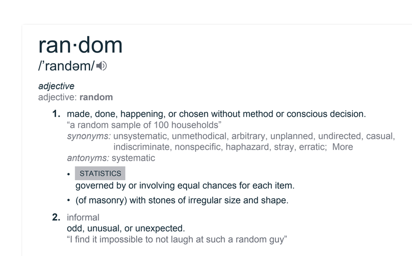


Exhibit 3.
Components of uncertainty


ran·dom
/ˈrandəm/
adjective
adjective: **random**
1. made, done, happening, or chosen without method or conscious decision.
“a random sample of 100 households”
synonyms: unsystematic, unmethodical, arbitrary, unplanned, undirected, casual, indiscriminate, nonspecific, haphazard, stray, erratic; More
antonyms: systematic
- STATISTICS
governed by or involving equal chances for each item.
- (of masonry) with stones of irregular size and shape.
2. informal
odd, unusual, or unexpected.
“I find it impossible to not laugh at such a random guy”

Random sampling
&

- **Random questions:** due to insufficiently explicit specification of the analytic question and assumptions.
- **Random data sources and quality:** due to variability among data sources, with variation in quality of measurement and information.
- **Random analysis:** due to data-driven model specification.
- **Random interpretation:** due to insufficiently explicit communication of the calculus of evidence.

Random understanding due to uncertainty beyond sampling error

DISCUSSION

Numerous threats militate against integrity of inference and the self-correcting mechanism of the scientific method. The credibility of ODA for scientific evidence is a prevalent concern.[15-17] Sampling error is only one component in a long sequence of distortive forces leading to clinical knowledge, and is often not the most important. Many other methodologic and systematic pitfalls introduce loss and distortion of the ‘signal’ representing causal relationships in nature, much as noise compromises signal in a communication channel.

Several systems have been promoted to evaluate the quality of ODA.[18-20] These either borrow from the RCT paradigm or address particular aspects of ODA. No overarching model providing a coherent and comprehensive approach to optimizing inferential integrity is widely acknowledged. There is similarity between transmission of an ‘image’ through the various filters and lenses of the ‘episcope,’[21] but advances in analytic methodology (Harrell[7,8]) and attention to the process and methods of replication and evidence valuation need to be included.

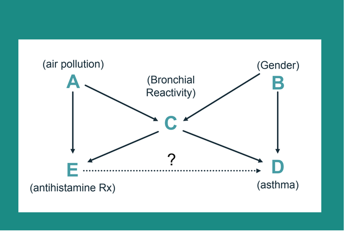
This perspective connects multiple-bias modeling, causal graph theory, modeling strategies, reproducible research methods and Bayesian epistemology for a coherent approach to ODA to optimizes inferential integrity. While not an exhaustive and sufficient set of criteria, they represent necessary components frequently overlooked in less circumspect approaches. These criteria are proposed not as a set of rules that determine study qualification, but as values that influence it.

CONCLUSIONS

The value of ODA for clinical decision making is constrained by uncertainties above and beyond those represented in estimated standard errors. Various authors and methodologists have made important contributions to various aspects of the problem. Integrating these into a methodologic thesis suggests guidelines for ODA practice and ODA evaluation. This model for ODA encourages a coherent approach focused on support for the self-correcting mechanism in science and optimization of inferential integrity.

1. Grounded in an explicit and detailed theory

- Explicit specification of the causal proposition and of all underlying causal and modeling assumptions is required for (a) correct and unbiased analysis, (b) reproducibility, (c) effective critical review of methods, (d) correct inference, (e) generalizability of results, (f) combining evidence generated by the study with evidence external to the study (evidence calculus) to advance knowledge.
- Structural models[2] and directed acyclic graphics (DAGs)[3-6] provide a unified framework for:
  - deducing statistical associations implied by causal relations
  - evaluating design and analysis strategies for any causal question
  - exposing all relevant considerations for critical review
  - efficiently communicating comprehensive, complex and nuanced information.



2. Regression modeling methods address uncertainties due to model specification, over-fitting, predictive optimism

- Important to develop and validate accurate models whose accuracy will not degrade when applied in external data sources.
- Accurate estimation depends on not over-fitting the adjustment variables. The unreliability of estimates equals the degree to which the overall model fails to validate on an independent sample.[7,8]
- The uncertainties due to model specification, over-fitting, predictive optimism should be a routine, formal and explicit component of reported results.

3. Methods for fully reproducible research were employed

- The introspective and *ad hoc* nature of the design of analyses frequently eludes any meaningful objective assessment of their performance.
- Reluctance to share data has been found to be associated with weaker evidence and a higher prevalence of apparent errors in the reporting of statistical results.[9]
- Methods for reproducible research are being formalized.[10,11] Incorporating both reproducibility, *and* the expectation for reproduction of research, into ODA increases transparency, facilitates effective critical review and reflects a commitment to the self-correcting mechanisms in science and scientific integrity.

Exhibit 4.
Criteria for ODA quality

1. Grounded in an explicit and detailed theory
- **Assumptions formally and exhaustively expressed in Directed Acyclic Graphics (DAGs)**
- **DAGs used to identify minimal set of adjusters for confounding and to rule out collider bias**

2. Regression modeling methods address uncertainties due to model specification, over-fitting, predictive optimism

3. Methods for fully reproducible research were employed

4. Quantitative bias analysis performed to evaluate the uncertainty due to residual bias

5. The finding was replicated in independent data

6. Expression and evaluation of results in terms of reduction of uncertainty and altering prior belief

4. Quantitative bias analysis performed to evaluate the uncertainty due to residual bias

- Conventional analytic results do not reflect any source of uncertainty other than random error. Standard errors cannot be expected to show us the indeterminacies and uncertainties we face.[12]
- The illusory precision of conventional results is rarely addressed by more than intuitive judgments based on flawed heuristics. An assessment of uncertainty due to assumptions (uncertainty analysis) is an essential part of inference.
- Multiple-bias modeling (Greenland[12]) should become a standard part of ODA.

5. The finding was replicated in independent data

- Tight feedback loops are important to the scientific method and the self-correcting mechanism of science.
- Better predictions are found in domains where tight feedback loops exist to test or corroborate hypotheses.[13]
- In engineering of communication systems, redundancy is the main tool for error reduction and signal recovery.[1] Similarly, reproduction of research and replication of findings are the main process for the self-correcting mechanism of science.
- Replication updates priors from evidence external to the study.

6. Expression and evaluation of results in terms of reduction of uncertainty and altering prior belief

- Competing accounts of causal relationships are not easily reconciled due to subjective conditioning of evidence and intuitive judgments.
- Inefficiencies accrue from not explicitly and quantitatively characterizing existing evidence prior to conducting a study and systematically representing how evidence should be revised with new data.
- Consensus building and evidence updating to advance knowledge are more effectively managed by combining evidence generated by ODA with evidence external to the study (evidence calculus).[14]

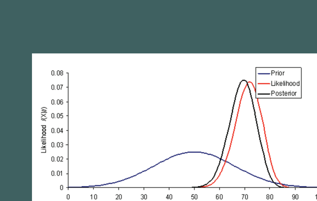
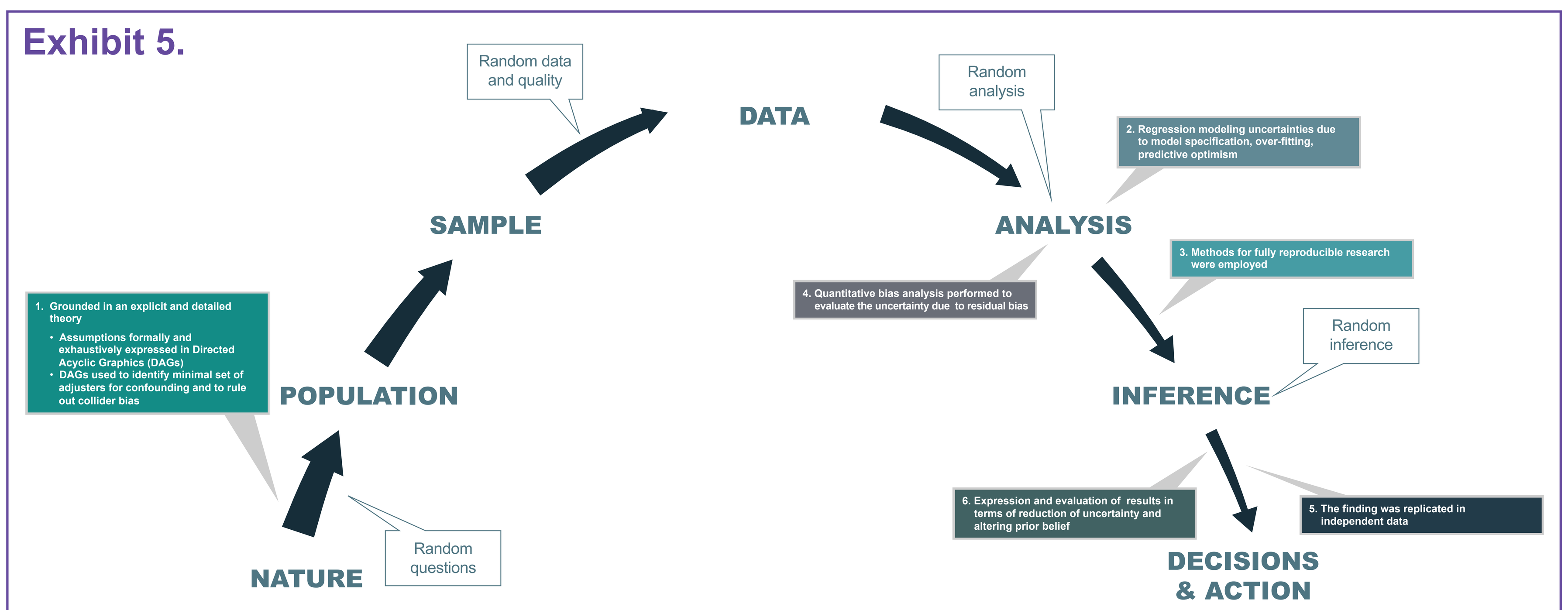


Exhibit 5.



REFERENCES

1. Shannon CE. A mathematical theory of communication. *Bell Syst Tech J*. 1948;27:379–423, 623–656.
2. Pearl J. *Causality: Models, Reasoning and Inference*. New York, NY: Cambridge University Press; 2009.
3. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology*. 1999;10:37–48.
4. Greenland S, Brumback B. An overview of relations among causal modelling methods. *Int J Epidemiol*. 2002;31:1030–1037.
5. VanderWeele TJ, Hernán MA, Robins JM. Causal directed acyclic graphs and the direction of unmeasured confounding bias. *Epidemiology*. 2008;19:720–728.
6. Greenland S. Quantifying biases in causal models: classical confounding vs collider-stratification bias. *Epidemiology*. 2003;14:300–306.
7. Harrell FE. *Regression Modeling Strategies With Applications to Linear Models, Logistic Regression, and Survival Analysis*. New York, NY: Springer; 2001.
8. Steyerberg EW. *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*. New York, NY: Springer; 2009.
9. Wicherts JM, Bakker M, Molenaar D. Willingness to share research data is related to the strength of the evidence and the quality of reporting of statistical results. *PLoS One*. 2011;6:e26828.
10. Xie Y. Knitr: a comprehensive tool for reproducible research in R. In: Stodden V, Leisch F, Peng RD, eds. *Implementing Reproducible Computational Research*. Boca Raton, FL: CRC Press; 2014:3–32.
11. Gandrud C. *Reproducible Research With R and RStudio*. Boca Raton, FL: CRC Press; 2014.
12. Greenland S. Multiple-bias modelling for analysis of observational data. *R Statist Soc A*. 2005;168(part 2):267–306.
13. Silver N. *The Signal and the Noise: Why Most Predictions Fail*. New York, NY: Penguin Group; 2012.
14. Goodman SN. Introduction to Bayesian methods I: measuring the strength of evidence. *Clin Trials*. 2005;2:282–290.
15. Unreliable research: trouble at the lab. *Economist*. 2013;409:26–30.
16. Singh AR, Singh SA. Guidelines, editors, pharma and the biological paradigm shift. *Mens Sana Monogr*. 2007;5:27–30.
17. Ioannidis JP. Why most published research findings are false. *PLoS Med*. 2005;2:e124.
18. Madigan D, Stang PE, Berlin JA, et al. A systematic statistical approach to evaluating evidence from observational studies. *Ann Rev Stat Its Application*. 2014;1:11–39.
19. Vandenbroucke JP, von Elm E, Altman DG, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *Epidemiology*. 2007;18:805–835.
20. Sauerbrei W, Abrahamowicz M, Altman DG, le Cessie S, Carpenter J; on behalf of the STRATOS initiative. STRengthening Analytical Thinking for Observational Studies: the STRATOS initiative. *Stat Med*. doi: 10.1002/sim.6265.
21. Maclure M, Schneeweiss S. Causation of bias: the episcope. *Epidemiology*. 2001;12:114–122.

International Conference on Pharmacoepidemiology & Therapeutic Risk Management; October 24-27, 2014; Taipei, Taiwan