

# Causation in Population Health Informatics and Data Science

Olaf Dammann • Benjamin Smart

# Causation in Population Health Informatics and Data Science

 Springer

Olaf Dammann  
Department of Public Health and  
Community Medicine  
Tufts University School of Medicine  
Boston, MA  
USA

Benjamin Smart  
The African Centre for Epistemology  
and Philosophy of Science  
University of Johannesburg  
Johannesburg  
South Africa

ISBN 978-3-319-96306-8      ISBN 978-3-319-96307-5 (eBook)  
<https://doi.org/10.1007/978-3-319-96307-5>

Library of Congress Control Number: 2018958761

© Springer Nature Switzerland AG 2019

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG  
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

# Preface

Human civilization as we know it would look very different were it not for collective efforts to maintain and improve population health. Access to the means that support individual and community health is a human right. The continued improvement of global public health must be an administrative and political priority, at local, regional, and national levels. So many health administrators, scientists, field workers, students, project coordinators, nurses, and physicians work tirelessly toward that goal. Better population health means better lives.

Academic schools and departments of medicine and public health are charged with the generation of new knowledge to help with this endeavor. Because effective medical and public health practice needs a solid scientific foundation, we need to devise methods that ensure the generation of unbiased data, reliable information, solid evidence, and useful knowledge. Most of these methods are computer-based; informatics and data science thus play a crucial role in this research. Accordingly, considerable attention has been paid over the past decade to the development of public health informatics, with the second edition of a comprehensive textbook [1] and the creation of an academic journal, the *Online Journal of Public Health Informatics* ([ojphi.org](http://ojphi.org)). Most recently, *Population Health Informatics* has been added to the mix [2]. The field covers a broad variety of informatics applications in population health research and practice. One crucial issue, however, is the lack of substantial effort to improve the *interpretation* of population health data and inference, i.e., the generation of new knowledge [3, p. 100–1]. The central issue here is the lack of methods for interpretation and inference that help turn data into information, information into evidence, and evidence into knowledge.

This is the void this book attempts to fill, by bringing in a third perspective that comes from the growing interest among philosophers in causal explanation and mechanisms in the health sciences. Imagine a Venn diagram with three circles: one is informatics and data science, the second is the health sciences, and the third is philosophy of science. The topic of this book is situated right in the middle of the multidisciplinary space defined by the diagram, where all three circles overlap and where, as of yet, not much has been published.

This book had a lengthy gestation and many have helped along the way. Olaf Dammann would like to thank Guanglan Zhang, Naftali Weinberger, Olaf Wolkenhauer, Alex Broadbent, Ted Poston, Paul Thagard, and all his students and colleagues at Tufts for many interesting and helpful discussions. First and foremost though, a big thanks to Christiane, Lina, and Laura for their love, patience, and support. Benjamin Smart would like to thank Veli Mitova and Michael Talibard for their comments on earlier versions of Chaps. 3 and 4, Em Watson for her love and support, and his coauthor for relentless patience. Both authors appreciate the continued guidance offered by our editor Grant Weston at Springer.

We hope to generate a fruitful debate that will develop some of the ideas discussed in this book. We have tried to offer food for thought in a palatable way, and all errors are ours. Our last chapter is a standing invitation to contact us to join the conversation. As the chef of a good restaurant would say, if you don't like our food, tell us; if you do, tell others.

## References

1. Magnuson JA, Fu PC. Public health informatics and information systems, Health informatics series. New York: Springer; 2014.
2. Joshi A, Thorpe L, Waldron L. Population health informatics: driving evidence-based solutions into practice. Burlington, Massachusetts: Jones & Bartlett Learning; 2019.
3. Tulchinsky TH, Varavikova E. The new public health. 2nd ed. Amsterdam/Boston: Elsevier/Academic Press; 2009. xxiv, 672 p.

Boston, MA, USA  
Johannesburg, South Africa

Olaf Dammann  
Benjamin Smart

# Contents

<b>1</b>	<b>Introduction</b> . . . . .	1
1.1	Background . . . . .	2
1.2	Etiology: Risk and Causation. . . . .	6
1.3	Causal Theory in Epidemiology . . . . .	8
1.4	Kantian Lament . . . . .	12
1.5	What This Book Is, and Is Not About . . . . .	13
1.6	Overview of Chapters . . . . .	14
<b>2</b>	<b>Health Data Science</b> . . . . .	15
2.1	Data Science. . . . .	15
2.2	The Technology Domain . . . . .	17
2.2.1	Computer Science . . . . .	17
2.2.2	Health Informatics . . . . .	18
2.2.3	Public Health Informatics . . . . .	18
2.3	The Analytics Domain . . . . .	21
2.3.1	Biostatistics . . . . .	21
2.3.2	Bioinformatics . . . . .	21
2.3.3	Epidemiology. . . . .	22
2.3.4	Simulation . . . . .	23
2.4	The Conceptual Domain . . . . .	23
2.4.1	Philosophy of Information . . . . .	23
2.4.2	Philosophy of Health and Causation . . . . .	24
2.5	Summary . . . . .	26
<b>3</b>	<b>The Metaphysics of Illness Causation</b> . . . . .	27
3.1	Introduction . . . . .	27
3.2	Illness Causation: Causes Plus Mechanisms . . . . .	28
3.2.1	Hume and the Regularity Accounts of Causation . . . . .	29
3.2.2	Rothman’s Pies and Mackie’s <i>INUS</i> Conditions . . . . .	31
3.2.3	The Counterfactual Account of Causation. . . . .	34
3.2.4	Dispositions: The Powers of Disease. . . . .	36
3.3	Summary . . . . .	39

<b>4</b>	<b>Causal Inference in Population Health Informatics</b> . . . . .	43
4.1	Introduction . . . . .	43
4.2	Causal Judgements. . . . .	44
4.2.1	Three Fundamental Problems . . . . .	44
4.2.2	Explanation and Medical Knowledge . . . . .	45
4.3	Knowledge . . . . .	46
4.3.1	The Problem of Induction . . . . .	46
4.3.2	Solutions to the Problem of Induction . . . . .	49
4.4	Epidemiology and Identifying Causation . . . . .	51
4.4.1	The Potential Outcomes Approach Revisited . . . . .	51
4.4.2	Observational Studies and Randomised Controlled Trials . . . . .	52
4.5	Hill's Heuristics . . . . .	52
4.5.1	Association or Causation? . . . . .	53
4.5.2	Hill Discussed . . . . .	53
4.5.3	Debunking Soft Criterianism . . . . .	54
4.6	Gold Standard?. . . . .	57
4.7	Is There Such a Thing as Medical Knowledge?. . . . .	59
4.7.1	The Tripartite Account of Knowledge . . . . .	59
4.7.2	Internalism, Externalism, and the KK Principle . . . . .	60
4.7.3	Externalism to the Rescue: The Causal Theory of Knowing . . . . .	60
4.8	Wrapping Up . . . . .	61
<b>5</b>	<b>Making Population Health Knowledge</b> . . . . .	63
5.1	Introduction . . . . .	63
5.2	The Knowledge Pyramid . . . . .	64
5.2.1	Ackoff's Hierarchy . . . . .	64
5.2.2	Wisdom or Knowledge?. . . . .	65
5.3	DIEK: Data, Information, Evidence, Knowledge . . . . .	66
5.3.1	Data Are Data Are Data . . . . .	67
5.3.2	Information Is Data Contextualized. . . . .	68
5.3.3	Evidence: Information Compared . . . . .	69
5.3.4	Knowledge from Evidence. . . . .	70
5.3.5	Preventive Intervention: Knowledge-Maker or Knowledge-Confirmer?. . . . .	74
5.4	Quality Transition Checkpoints . . . . .	75
5.5	Summary . . . . .	77
<b>6</b>	<b>Population Risk</b> . . . . .	79
6.1	Introduction . . . . .	79
6.2	Prediction . . . . .	80
6.2.1	Individual Risk. . . . .	81
6.2.2	Population Risk . . . . .	82
6.2.3	Population Risk Estimates Used in Individuals . . . . .	83
6.2.4	Why Is All This Problematic? . . . . .	85

- 6.3 Population Risk Estimation . . . . . 87
  - 6.3.1 Univariable Risk Analysis . . . . . 88
  - 6.3.2 Multivariable Risk Estimation . . . . . 91
- 6.4 Risk Estimation as Causal Explanation . . . . . 94
  - 6.4.1 Risk Factors and Causes . . . . . 95
  - 6.4.2 Randomized-Controlled Trials as Cause-Makers . . . . . 96
- 6.5 Summary . . . . . 98
- 7 Integrating Evidence.** . . . . . 99
  - 7.1 Introduction . . . . . 99
    - 7.1.1 Evidence in Public Health . . . . . 100
  - 7.2 Target . . . . . 102
  - 7.3 Method . . . . . 102
    - 7.3.1 Systems Health. . . . . 102
    - 7.3.2 Systems Biology . . . . . 104
    - 7.3.3 Systems Medicine . . . . . 105
    - 7.3.4 Systems Epidemiology . . . . . 105
  - 7.4 Manipulation . . . . . 109
  - 7.5 Complexity and Simplicity . . . . . 112
    - 7.5.1 Complexity. . . . . 112
    - 7.5.2 Simplicity. . . . . 114
  - 7.6 Summary . . . . . 115
- 8 Conclusion and Invite.** . . . . . 117
- References** . . . . . 119
- Index.** . . . . . 131