

SPRINGER BRIEFS IN PHILOSOPHY

Veli-Pekka Parkkinen · Christian Wallmann  
Michael Wilde · Brendan Clarke  
Phyllis Illari · Michael P. Kelly  
Charles Norell · Federica Russo · Beth Shaw  
Jon Williamson

# Evaluating Evidence of Mechanisms in Medicine

## Principles and Procedures



Springer Open

# **SpringerBriefs in Philosophy**

SpringerBriefs present concise summaries of cutting-edge research and practical applications across a wide spectrum of fields. Featuring compact volumes of 50 to 125 pages, the series covers a range of content from professional to academic. Typical topics might include:

- A timely report of state-of-the art analytical techniques
- A bridge between new research results, as published in journal articles, and a contextual literature review
- A snapshot of a hot or emerging topic
- An in-depth case study or clinical example
- A presentation of core concepts that students must understand in order to make independent contributions

SpringerBriefs in Philosophy cover a broad range of philosophical fields including: Philosophy of Science, Logic, Non-Western Thinking and Western Philosophy. We also consider biographies, full or partial, of key thinkers and pioneers.

SpringerBriefs are characterized by fast, global electronic dissemination, standard publishing contracts, standardized manuscript preparation and formatting guidelines, and expedited production schedules. Both solicited and unsolicited manuscripts are considered for publication in the SpringerBriefs in Philosophy series. Potential authors are warmly invited to complete and submit the Briefs Author Proposal form. All projects will be submitted to editorial review by external advisors.

SpringerBriefs are characterized by expedited production schedules with the aim for publication 8 to 12 weeks after acceptance and fast, global electronic dissemination through our online platform SpringerLink. The standard concise author contracts guarantee that

- an individual ISBN is assigned to each manuscript
- each manuscript is copyrighted in the name of the author
- the author retains the right to post the pre-publication version on his/her website or that of his/her institution.

More information about this series at <http://www.springer.com/series/10082>

Veli-Pekka Parkkinen · Christian Wallmann  
Michael Wilde · Brendan Clarke  
Phyllis Illari · Michael P. Kelly  
Charles Norell · Federica Russo  
Beth Shaw · Jon Williamson

# Evaluating Evidence of Mechanisms in Medicine

Principles and Procedures

Veli-Pekka Parkkinen  
Department of Philosophy  
University of Bergen  
Bergen, Norway

Christian Wallmann  
Centre for Reasoning  
University of Kent  
Canterbury, UK

Michael Wilde  
Centre for Reasoning  
University of Kent  
Canterbury, UK

Brendan Clarke  
Department of Science  
and Technology Studies  
University College London  
London, UK

Phyllis Illari  
Department of Science  
and Technology Studies  
University College London  
London, UK

Michael P. Kelly  
Public Health and Primary Care  
University of Cambridge  
Cambridge, UK

Charles Norell  
Cancer Research UK  
London, UK

Federica Russo  
Department of Philosophy  
University of Amsterdam  
Amsterdam, Noord-Holland  
The Netherlands

Beth Shaw  
Centre for Evidence-Based Policy  
Oregon Health & Science University  
Portland, OR, USA

Jon Williamson  
Centre for Reasoning  
University of Kent  
Canterbury, UK



ISSN 2211-4548

SpringerBriefs in Philosophy

ISBN 978-3-319-94609-2

<https://doi.org/10.1007/978-3-319-94610-8>

ISSN 2211-4556 (electronic)

ISBN 978-3-319-94610-8 (eBook)

Library of Congress Control Number: 2018945456

© The Editor(s) (if applicable) and The Author(s) 2018. This book is an open access publication.

**Open Access** This book is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this book are included in the book's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the book's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

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.

Printed on acid-free paper

This Springer imprint is published by the registered company Springer International Publishing AG part of Springer Nature

The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

*To all our mentors and all our students.*

# Foreword

Techniques designed to evaluate the evidence of efficacy, effectiveness, and safety in medicine abound. They are almost invariably based on the results of experimental studies (primarily randomised controlled trials) or observational studies (such as case-control designs). None, so far as I am aware, include evidence of mechanisms.

This is perhaps surprising given the views of Sir Austin Bradford Hill. In his famous Presidential Address to the Section of Occupational Health of the Royal Society of Medicine in 1965, he outlined nine factors that should be ‘taken into account’ when deciding whether an ‘association’ was ‘causal’. One of these factors was what he called ‘biological plausibility.’ Yet, despite the growth of the evidence-based medicine movement—many of whose principles have their genesis in Bradford Hill’s famous lecture—none (in so far as I am aware) have included evidence of mechanisms as part of their approach.

This work is therefore not just a timely reminder of the importance of mechanisms. It is also a wake-up call to the evidence-based medicine movement to incorporate mechanisms in their evaluation of ‘evidence.’ EBM+ comes of age.

London, UK  
April 2018

Professor Sir Michael D. Rawlins

# Preface

The themes explored in this book began with a paper written by two of the authors (Russo and Williamson 2007), which set out the core idea of evidential pluralism in the context of establishing causation in medicine. An exploratory grant from the UK Arts and Humanities Research Council for the project *Mechanisms and the evidence hierarchy* allowed Brendan Clarke, Donald Gillies, Phyllis Illari, Federica Russo, and Jon Williamson to develop a collaboration via a series of informal workshops. This collaboration led to the publication of two papers which developed these themes of evidence and mechanisms (Clarke et al. 2013, 2014).

This book was written during the course of two further three-year research projects on evidence of mechanisms, connected to the EBM+ consortium. EBM+ aims to be a hub for research that contributes to our understanding of the role of evidence of mechanisms in medical methodology. The research projects in question were the Leverhulme-funded project *Grading evidence of mechanisms in physics and biology*, which involved Stefan Dragulinescu, Veli-Pekka Parkkinen, and Jon Williamson, and the AHRC-funded project *Evaluating Evidence in Medicine*. This latter project involved Brendan Clarke, Athena Drakou, Donald Gillies, Phyllis Illari, Mike Kelly, Charles Norell, Federica Russo, Beth Shaw, Kurt Straif, Jan Vandenbrouke, Christian Wallmann, Michael Wilde, and Jon Williamson.

More widely, this work benefited from numerous interactions between scientists, evidence appraisal practitioners and philosophers. These collaborations were sustained by a constant effort, on all sides, to translate jargon and to explain domain-specific problems and priorities. Inter- and transdisciplinary translation is a difficult exercise, and we greatly appreciate the dedication, open-mindedness, and patience of those with whom we have interacted. Our long-term project is to contribute to the preparation of guidance in various areas of medical practice, by providing conceptual tools that can add to current evaluation instruments. This requires, more generally, addressing philosophical problems and challenges as they arise in the practice of medicine. We would be keen to hear from others with similar interests.

**The aim of this book.** The aim of this book is twofold. On the one hand, we develop an approach to evidence evaluation that complements the existing methods used in EBM (and evidence-based policy more generally) by explicating the role of evidence of mechanisms when assessing causal claims in medicine. On the other hand, we aim to contribute to existing philosophical debates about evidence in medicine by giving a detailed account of how evidence of mechanisms can be evaluated.

**Who should read this book.** This book is intended for those who are interested in evidence in the health sciences. This includes those who work directly with evidence, such as guideline developers and those charged with evidence appraisal. We are also writing for those whose interest in evidence is more conceptual. This includes philosophers of science and medicine, as well as those who produce or interpret guidance on the effectiveness of health interventions. This latter group might include policy-makers, journalists, and politicians.

**How to use this book.** This book was written for those interested in philosophical and practical questions that arise during the use of evidence in medicine, public health, and social care. This is an extremely broad potential audience, and the parts of this book approach these concepts in several different ways. Most of the applied material is concentrated in Parts II and IV. Part II includes a variety of tools for working with evidence of mechanisms suitable for different contexts, while Part IV presents some specific applications of the ideas presented in the book. More theoretical material can be found in Part I and Part III.

We have identified several likely paths that readers with different interests might most fruitfully engage with the issues raised in this book.

For *clinical practitioners*, we would recommend that you begin by looking at Part I in order to gain background information about mechanisms and evidence of mechanisms. We would then suggest moving on to Part II. The tools presented in Part II provide a way of applying the strategies developed in this book directly to commonly encountered procedures in evidence appraisal, and they have been developed so that they can be used independently of the more theoretical parts of the book. They will provide a foundation for reading the more theoretical parts of the book.

For *policy-makers, guideline developers, and others involved in interpreting evidence in the policy context*, we would recommend reading Part I, before moving on to the tools from Part II. Beginning in this way should leave the reader confident to navigate the rest of the material here. We have also provided a series of particular applications (in Part IV) which contain material of possible interest.

For *philosophers of medicine*, we would suggest progressing from Part I to the more theoretical parts of the book (i.e. Part III), and then returning to the more applied material in the other chapters.

Appendices and a glossary are available at [ebmplus.org/appendices](http://ebmplus.org/appendices).

Bergen, Norway

Canterbury, UK

Canterbury, UK

London, UK

London, UK

Cambridge, UK

London, UK

Amsterdam, The Netherlands

Portland, USA

Canterbury, UK

April 2018

Veli-Pekka Parkkinen

Christian Wallmann

Michael Wilde

Brendan Clarke

Phyllis Illari

Michael P. Kelly

Charles Norell

Federica Russo

Beth Shaw

Jon Williamson

## References

- Clarke, B., Gillies, D., Illari, P., Russo, F., & Williamson, J. (2013). The evidence that evidence-based medicine omits. *Preventive Medicine*, *57*(6), 745–747.
- Clarke, B., Gillies, D., Illari, P., Russo, F., & Williamson, J. (2014). Mechanisms and the evidence hierarchy. *Topoi*, *33*(2), 339–360.
- Russo, F., & Williamson, J. (2007). Interpreting causality in the health sciences. *International Studies in the Philosophy of Science*, *21*(2), 157–170.

# Acknowledgements

This research was supported by a grant from the UK Arts and Humanities Research Council for a project *Evaluating evidence in medicine* and a grant from the Leverhulme Trust for a project *Grading evidence of mechanisms in physics and biology*. The research benefited greatly from the following organisations allowing us to observe evidence appraisal meetings: the International Agency for Research on Cancer (IARC), the UK Medicines and Healthcare Products Regulatory Agency (MHRA), and the UK National Institute for Health and Care Excellence (NICE). The core project team includes the authors of this monograph together with Kurt Straif, head of monographs at IARC, Donald Gillies, emeritus professor at University College London, and Jan Vandenbroucke, emeritus professor at Leiden University, to whom we are very grateful for many helpful discussions. We are very grateful to these latter members of the team for their contribution to the ideas of this book and for their comments on earlier versions of this text. We are also very grateful to other members of the EBM+ network ([ebmplus.org](http://ebmplus.org)) for valuable discussions, to Nancy Cartwright for comments on a previous manuscript, to an anonymous reviewer for Springer, to UCL Health Creatives for designing Figs. 4.1 and 4.2 and Tables 4.1–4.8, and to Silvie Demandt and Lucy Fleet at Springer.

This book is a collaborative work. Every author read and commented on the whole manuscript and contributed significantly to its writing. We divided up the task of writing first drafts of specific sections, which were subsequently submitted to the whole group. In particular: Chap. 1 was first drafted by Brendan Clarke; Chaps. 2 and 3 by Jon Williamson; Chap. 4 by Brendan Clarke, Phyllis Illari, Mike Kelly, Charles Norrell, Veli-Pekka Parkkinen, Federica Russo and Beth Shaw; Chap. 5 by Michael Wilde apart from Sect. 5.6 which was developed by Beth Shaw; Chap. 6 by Veli-Pekka Parkkinen; Chap. 7 by Christian Wallmann; Chap. 8 by Veli-Pekka Parkkinen, Christian Wallmann, Michael Wilde and Jon Williamson; Chap. 9 by Federica Russo and Mike Kelly; and Chap. 10 by Christian Wallmann. Jon Williamson led the development of the theoretical approach and coordinated the writing of the book. Brendan Clarke led the editing and harmonising of all the sections in the final stage of writing.

# Contents

## Part I Why Consider Mechanisms?

<b>1</b>	<b>Introduction</b> . . . . .	3
1.1	What is the Key Evidence? . . . . .	4
1.2	The Process of Evaluating Evidence . . . . .	6
1.3	Our Approach to Evaluating Evidence . . . . .	8
	References . . . . .	8
<b>2</b>	<b>An Introduction to Mechanisms</b> . . . . .	11
2.1	Mechanisms at a Glance . . . . .	11
2.2	What is a Mechanism? . . . . .	13
2.3	Why Consider Evidence of Mechanisms? . . . . .	15
	References . . . . .	20
<b>3</b>	<b>How to Consider Evidence of Mechanisms: An Overview</b> . . . . .	23
3.1	Questions to Address . . . . .	23
3.2	Quality of Evidence and Status of Claim . . . . .	25
3.3	Overall Approach . . . . .	27
	References . . . . .	33

## Part II Tools for Working with Mechanisms

<b>4</b>	<b>Tools</b> . . . . .	37
4.1	Introduction . . . . .	37
4.2	Is Your Policy Really Evidence-Based? . . . . .	39
4.3	Mechanisms in Clinical Research Appraisal Tool . . . . .	42
4.4	Mechanisms in Basic Science Research Appraisal Tool . . . . .	47
4.5	Critical Appraisal Tool for Evidence of Mechanisms . . . . .	50
4.6	GRADE-Style Tables for Mechanism Assessment . . . . .	50
4.7	Public Health and Social Care Tool . . . . .	56
	References . . . . .	58

### Part III Core Principles

<b>5</b>	<b>Gathering Evidence of Mechanisms</b> . . . . .	63
5.1	Identify Specific Mechanism Hypotheses . . . . .	64
5.2	Formulate the Review Questions . . . . .	68
5.3	Search the Literature . . . . .	69
5.4	Refine Results of the Search . . . . .	70
5.5	Presenting the Evidence of Mechanisms . . . . .	71
5.6	Worked Example on Probiotics and Dental Caries . . . . .	73
	References . . . . .	74
<b>6</b>	<b>Evaluating Evidence of Mechanisms</b> . . . . .	77
6.1	Overview . . . . .	77
6.2	Evaluating Mechanistic Studies . . . . .	80
6.3	Determining the Status of the General Mechanistic Claim . . . . .	82
6.4	Presenting the Quality of Evidence of Mechanisms . . . . .	85
	References . . . . .	89
<b>7</b>	<b>Using Evidence of Mechanisms to Evaluate Efficacy and External Validity</b> . . . . .	91
7.1	Efficacy . . . . .	91
7.2	External Validity . . . . .	94
7.3	Presenting the Status of a Causal Claim . . . . .	97
	References . . . . .	98

### Part IV Particular Applications

<b>8</b>	<b>Assessing Exposures</b> . . . . .	101
8.1	Comparison to IARC . . . . .	102
8.2	Comparison to SYRINA . . . . .	107
	References . . . . .	109
<b>9</b>	<b>Assessing Mechanisms in Public Health</b> . . . . .	111
9.1	Introduction . . . . .	111
9.2	Public Health and Evidence-Based Medicine in the UK . . . . .	112
9.3	Statistical Associations and Correlations in Public Health . . . . .	112
9.4	Recurrent Public Health Problems—Non-communicable Disease in the Present . . . . .	113
9.5	The Individual Level and the Population Level . . . . .	114
9.6	The Biological Level and the Social Level . . . . .	116
9.7	Mechanisms of Disease and Mechanisms of Prevention . . . . .	116
	References . . . . .	119
<b>10</b>	<b>Particularisation to an Individual</b> . . . . .	121
	References . . . . .	125

# Abstract

The use of evidence in medicine is something we should continuously seek to improve. This book seeks to develop our understanding of evidence of mechanism in evaluating evidence in medicine, public health, and social care; and also offers tools to help implement improved assessment of evidence of mechanism in practice. In this way, the book offers a bridge between more theoretical and conceptual insights and worries about evidence of mechanism and practical means to fit the results into evidence assessment procedures.

The book is designed so that the reader can use different parts, according to their primary aims.

Part I offers brief introductions to theoretical ideas developed in more depth in later chapters. It functions to orient the reader quickly with respect to the key issues: what evidence of mechanism is, the benefits of making its use more explicit, and the outline of the EBM+ approach to evidence of mechanism in evidence assessment.

Part II offers tools that can be used to improve the assessment of evidence of mechanism alongside evidence of correlation. Tools can be used in isolation or in the combinations suggested. The starting place is an overview tool, ‘Is your policy really evidence-based?’ Then further tools are provided for guideline developers for medical practice; a critical appraisal tool for politicians, journalists, academics, and so on; and a tool designed specifically for public health and social care.

Part III develops more theoretical ideas. It begins with the question of gathering evidence of mechanisms, addressing the issue that the relevant studies are not all indexed in the standard way of clinical trials (Chap. 5). Chapter 6 offers a process for evaluating evidence of mechanisms, by first breaking down into specific mechanism hypotheses, then combining the assessment into an evaluation of the quality of evidence for a general mechanism hypothesis. The part finishes in Chap. 7 by addressing how to integrate quality of evidence of the mechanism hypothesis with evidence of correlation, to come to an overall assessment of the quality of evidence for the causal claim.

Part IV examines some specific problems in applying evaluation of evidence of mechanisms to particular domains.

The whole book can be used, or those with a more practical focus can use Parts I and II, while those with a more theoretical interest can use Parts I and III, supplementing with chapters from Part IV as appropriate.

**Part I**  
**Why Consider Mechanisms?**

# Chapter 1

## Introduction



**Abstract** This chapter introduces the idea of EBM+, which adopts the explicit requirements of EBM, to (1) make all the key evidence explicit and (2) adopt explicit methods for evaluating that evidence. EBM+ then sets out to get us better causal knowledge by explicitly integrating evidence of mechanism alongside evidence of correlation. This chapter summarises some important benefits of including evidence of mechanism, particularly given how highly idealised study populations typically are, and introduces the need to make uses of evidence of mechanism more explicit.

This book describes a number of methods that integrate the appraisal of evidence of mechanisms with other forms of evidence. While these methods are relevant to many fields where evidence is assessed (see Clarke and Russo 2016), our starting point is evidence-based medicine (EBM). The methods in this book build upon the tools already developed by EBM, by taking evidence of mechanisms into account in addition to the clinical studies that are the focus of EBM. We refer to this development as EBM+.

**EBM+** Evidence of mechanisms should be integrated with evidence of correlation to better assess causal claims.

Medical practice depends fundamentally on the assessment of causal claims:

### Examples of assessing causal claims in medicine.

- Identifying the causes of cancers in humans.
- Evaluating whether a medical device will lead to improved outcomes in a particular patient.
- Establishing whether a public health action will have the desired effects in the target population.
- Determining whether a medicine has a specific detrimental side-effect.
- Ascertaining the cost effectiveness of a health intervention on a target population.

Causal claims underpin evidence-based medicine, guideline development, personalised medicine, narrative medicine, and other aspects of medicine.

This book concentrates on EBM because we explicitly endorse two core EBM principles:

**Two principles of EBM endorsed in this book.**

1. Make all the key evidence explicit.
2. Adopt explicit methods for evaluating that evidence.

These principles have been largely responsible for the significant advances made by EBM. In particular, EBM prompted the widespread adoption of techniques for analysing data on medical interventions, with the objective of determining whether these interventions are in fact delivering the expected results.

In this book, these principles are developed with respect to evidence of mechanisms. First, evidence of mechanisms is often key evidence and needs to be made explicit. Second, evidence of mechanisms needs to be explicitly evaluated when assessing a causal claim.

## 1.1 What is the Key Evidence?

EBM has hitherto focused primarily on one kind of evidence for a causal claim: evidence arising from clinical studies, including randomised trials and observational studies. However, this book is motivated by the idea that evidence for causal claims in medicine cannot simply be reduced to evidence of correlation. In the philosophy of causality, the following thesis has been put forward (Russo and Williamson 2007):

**Evidential pluralism.** This is the thesis that one typically needs both evidence of correlation and evidence of mechanisms to establish a causal claim.

Evidential pluralism is relevant to deciding what counts as key evidence. As we shall explain, the supposition that the key evidence will be all of one type (e.g., evidence from RCTs) is not a good one. Note that this thesis about forms of evidence goes beyond the (intuitively appealing) idea that taking more evidence into account will lead to better inferences.

To develop this argument, two pieces of terminology will be helpful: [efficacy](#) and [effectiveness](#). (Technical terms are hyperlinked to their definitions. A full glossary is available in the [online appendices](#).) Although these are likely to be familiar to most readers because of their widespread use in the medical literature, our usage of these terms is broader than their usual meaning. We define these terms as follows:

**Efficacy** concerns the effect(s) of some intervention or exposure in a particular study population. An *efficacy claim* is a claim that the intervention or exposure has some specific effect in the study population.

**Effectiveness** concerns the effect(s) of an intervention or exposure in some target population of interest, such as a population of patients to be treated. An *effectiveness claim* is a claim that the intervention or exposure has some specific effect in the target population.

The term ‘efficacy’ is normally only used in the context of a *beneficial* effect of an *intervention*. However, what we have to say in this book applies equally when assessing whether an intervention causes some particular *harm*, or when assessing whether an *exposure* causes a particular harm (or, indeed, benefit). So we use ‘efficacy’ throughout this book in a more general sense, covering harms as well as benefits and exposures as well as interventions. Similarly for ‘effectiveness’.

When a relationship applies more broadly than in a study population, it is sometimes said to be *externally valid*:

**External validity** concerns an inference from a study to a target population. If a causal claim that holds in a study population can be extrapolated to a target population of interest, then it may be described as externally valid.

To use the terminology of Cartwright and Hardie (2012, 15), external validity concerns how we go from *knowing that something works somewhere* (efficacy) to *knowing that it will work for us* (effectiveness). Extrapolation is typically crucial for demonstrating effectiveness:

**Effectiveness = efficacy + external validity.** Typically, one establishes that a causal claim holds in a target population by establishing the claim in a study population and then extrapolating that claim to the target population.

The reason for proceeding to effectiveness via efficacy and external validity is that a study population is typically highly idealised, and thus differs from the target population in important ways. For example, a study population for evaluating the effectiveness of a drug might exclude those with multiple morbidities or pregnant women; a study population for evaluating the carcinogenicity of an environmental exposure might be a laboratory population of an animal model. Establishing external validity is crucial because the mechanism of action in the study population may not be particularly robust.

An **idealised population** is one which is a non-representative subpopulation of the general population. Idealised populations satisfy certain ideal experimental conditions or experience a narrowly circumscribed range of exposures.

A **robust** mechanism is one that works in the same way across a wide variety of background conditions; a **fragile** mechanism does not.

As we shall see, evidence of mechanisms is crucial to establishing both efficacy and external validity. While evidence of mechanisms is already implicit in, for example, the design of clinical trials, mechanistic studies are generally not explicitly evaluated when making policy or treatment decisions (Clarke et al. 2013, 2014). This is largely a consequence of the downplaying of mechanistic studies in the most influential EBM methods manuals (such as GRADE), owing to concerns about possible bias. While we acknowledge that there are valid concerns about biases, we regard this wholesale downplaying as a mistake. At present, evidence of mechanisms does in fact influence the evaluation of effectiveness. For example, there may be evidence that the mechanism of action in a study population is rather different from those in a target population and this difference can be taken into account when assessing the effectiveness of a drug. But this influence of evidence of mechanisms is often invisible, because it is mediated by the opinions of experts, particularly expert panel members on evidence appraisal committees. This influence is reasonable: evidence of mechanisms plays a vital role in providing evidence of effectiveness. However, the lesson of evidence-based medicine is that one needs to make evidence explicit in order to scrutinise and challenge it properly, and that one needs to make explicit the ways in which evidence is evaluated in order to improve these methods of evaluation. This book seeks to extend this evidence-based approach to include evidence of mechanisms.

Evidence of mechanisms is often produced by means other than clinical studies. In philosophy of science, much attention has been devoted to the concept of mechanism in biology and medicine, as well as in many other scientific domains (see Chap. 2 for an introduction to mechanisms). However, comparatively less attention has been devoted to the question of how evidence of mechanisms is generated and assessed, especially in the context of medical practice. This gives rise to the next major theme of this book: how should we evaluate our evidence?

## 1.2 The Process of Evaluating Evidence

If—in common with many of those interested in EBM—your first exposure to the methods of EBM came from the profusion of introductory articles published in the medical literature in the late 1990s (such as Sackett et al. 1996), you might get the impression that the quality of a piece of clinical research could be determined

with relatively straightforward judgements of the methodology used in the research. Was the research randomised? Did the authors use intention-to-treat analysis? Had the statistical analysis produced a significant result? Unless these conditions were jointly satisfied, the research was of very low quality. And if they were satisfied, then it was likely that the work was of high quality, and should be used as a guide to practice—unless very serious provisos were detected (such as research misconduct).

This is because the evaluation of evidence in early EBM was about describing the methods used to produce that evidence. This placed the onus of judging the quality of a piece of research largely on the reader. In turn, this led to an emphasis on critiquing research methods as a proxy for judging the quality of research (Greenhalgh 2014, 28). Concerns about bias were given priority, and this heightened scrutiny of research methods has been the major defence against biased research.

However, critiquing research methods (rather than the details of a specified piece of research) is only possible because—for all the many complications of doing clinical research—many individual clinical studies share the same fundamental design. This means that shared ways of evaluating quality can be fairly easily learned and applied by health scientists, with the reasonable expectation that these simple methods are effective in stripping out biased research.

There is a fallacy here. Evaluating a small number of indicators did much of the work in downgrading biased research, and it did it in an efficient and simple way. Yet that is not to say that these techniques worked without any judgement on the part of the evaluator. Nor did these techniques work flawlessly. Although some research designs are more prone to bias than others, it does not automatically follow that, for instance, *all* non-randomised research is intrinsically biased. To use the terminology devised by Kahneman (2011), this kind of evaluation is a kind of system 1 thinking: fast and easy, but prone to faults. We are rightly suspicious of other kinds of system 1 thinking because of its propensity to bias. But sometimes speed is preferable to accuracy, and system 1 thinking may often be good enough. And we might choose to evaluate, for instance, clinical studies in this system 1 manner because there is a common structure of clinical research that allows us to make good enough judgements about their likely quality (Kahneman 2011). If EBM was a useful first approximation to evidence evaluation, then EBM+ is intended as a second, improved, approximation.

The same assumptions about commonality of methods do not seem to apply to evidence of mechanisms. It is hard to think of a field with more methodological diversity than contemporary bioscience research. For example, computer simulations, 31-P NMR, mass spectrometry, knockout studies and immunofluorescence do not exhaust the space of research strategies that have been used to understand a single protein (Mitchell and Dietrich 2006). And so we do not offer, in this book, a tool capable of evaluating all of this research in a substance-blind manner. We note in passing too that the presence of candidate indicators for clinical studies (such as intention to treat analysis, randomisation, or trial registration) that have been touted as ensuring that a piece of research can be accepted without question do no such thing, although they are individually helpful to an expert judge of clinical evidence. We need to judge evidence, and the methods and tools provided here are an aid to judgement, rather than a replacement for it.

### 1.3 Our Approach to Evaluating Evidence

The approach to evaluating evidence that is developed in this book can be traced back to work of Russo and Williamson (2007), who put forward an account of evidential pluralism in medicine. Williamson (2018) offers a recent defence of evidential pluralism.

Evidential pluralism in medicine is not a new idea. For instance, the causal indicators put forward by Hill (1965) can be viewed as a version of evidential pluralism. Several of Hill's indicators of causality are good indicators of mechanisms, while others are good indicators of correlation. We discuss Hill's indicators, and explain how our approach improves over them, in Chap. 6; see also Williamson (2018b).

The methods for evidence evaluation that we set out in later parts of the book all require judgement on the part of the user (Kelly and Moore 2012; Montgomery 2005). We do not pretend that there is a shibboleth or an algorithm that determines the excellence (or otherwise) of a piece of evidence of mechanisms. All evaluations of quality of evidence are fallible. With this work, we hope to reach those readers interested in combining practical methods for evidence evaluation with philosophical analysis.

## References

- Cartwright, N., & Hardie, J. (2012). *Evidence-based policy: A practical guide to doing it better*. Oxford University Press.
- Clarke, B., & Russo, F. (2016). Causation in medicine. In J. Marcum (Eds.), *Companion to contemporary philosophy of medicine* (pp. 297–322). Bloomsbury.
- Clarke, B., Gillies, D., Illari, P., Russo, F., & Williamson, J. (2013). The evidence that evidence-based medicine omits. *Preventive Medicine, 57*(6), 745–747.
- Clarke, B., Gillies, D., Illari, P., Russo, F., & Williamson, J. (2014). Mechanisms and the evidence hierarchy. *Topoi, 33*(2), 339–360.
- Greenhalgh, T. (2014). *How to read a paper: The basics of evidence-based medicine*. New York: Wiley.
- Hill, A. B. (1965). The environment and disease: Association or causation? *Proceedings of the Royal Society of Medicine, 58*, 295–300.
- Kahneman, D. (2011). *Thinking, fast and slow*. London: Penguin.
- Kelly, M. P., & Moore, T. A. (2012). The judgement process in evidence-based medicine and health technology assessment. *Social Theory and Health, 10*(1), 1–19.
- Mitchell, S. D., & Dietrich, M. R. (2006). Integration without unification: An argument for pluralism in the biological sciences. *The American Naturalist, 168*(S6), S73–S79.
- Montgomery, K. (2005). *How doctors think: Clinical judgment and the practice of medicine*. Oxford University Press.
- Russo, F., & Williamson, J. (2007). Interpreting causality in the health sciences. *International Studies in the Philosophy of Science, 21*(2), 157–170.

- Sackett, D. L., Rosenberg, W. M., Gray, J. M., Haynes, R. B., & Richardson, W. S. (1996). Evidence based medicine: What it is and what it isn't. *British Medical Journal*, 312, 71–2.
- Williamson, J. (2018a). Establishing causal claims in medicine. *International Studies in the Philosophy of Science*, in press. Preprint available at <http://blogs.kent.ac.uk/jonw/>.
- Williamson, J. (2018b). Establishing the teratogenicity of Zika and evaluating causal criteria. *Synthese*, in press. Preprint available at <http://blogs.kent.ac.uk/jonw/>.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



# Chapter 2

## An Introduction to Mechanisms



**Abstract** This chapter offers a brief summary of mechanisms, as including complex-system mechanisms (a complex arrangement of entities and activities, organised in such a way as to be regularly or predictably responsible for the phenomenon to be explained) and mechanistic processes (a spatio-temporal pathway along which certain features are propagated from the starting point to the end point). The chapter emphasises that EBM+ is concerned with evidence of mechanisms, not mere just-so stories, and summarises some key roles assessing evidence of mechanisms can play, particularly with respect to assessing efficacy and external validity.

This chapter introduces mechanisms and their use in the context of working with evidence in medicine. The first section gives an extremely short introduction to mechanisms that assumes no prior knowledge. Subsequent sections develop our account of mechanisms in more detail.

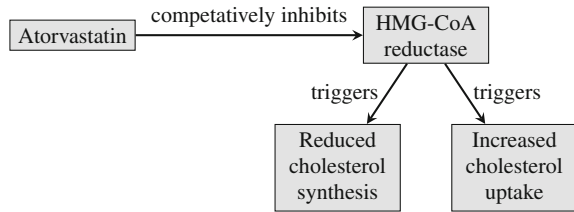
### 2.1 Mechanisms at a Glance

Mechanisms allow us to understand complex systems (e.g., physiological or social systems) and can help us to explain, predict, and intervene. An important subclass of mechanisms is characterised by the following working definition:

A complex-systems mechanism for a phenomenon consists of entities and activities organised in such a way that they are responsible for the phenomenon (Illari and Williamson 2012, 120).

For the example mechanism of Fig. 2.1, the phenomena are the effects of a drug, the drug and the receptor are the parts, and the interactions are the binding and triggering.

**Fig. 2.1** A simple example of a mechanism diagram that shows the effects produced by a drug via receptor binding



**Why do mechanisms matter?** Mechanisms explain how things work. This makes them important in their own right, but also means that they are often used when designing clinical studies. For example, one might decide to use a biomarker to evaluate the effect of a drug, and that decision would rely on our knowledge of some mechanism that links the biomarker with the drug. Note that while mechanisms of drug action are an important kind of mechanism, they are not the only kinds of mechanism that we will consider here.

**Caveats:**

- We will be interested in *evidence* of mechanisms, not descriptions of mechanisms for which there is no evidence. To be useful, descriptions of mechanisms should be connected to high-quality research, and not just to either background knowledge or to what Pawson (2003) calls ‘programme theories’. Otherwise they are merely just-so stories. Descriptions of mechanisms need to be supported by evidence to be useful.
- Mechanistic studies are not normally sufficient on their own to justify treatment or policy decisions. Other supporting evidence (such as that arising from clinical studies) is normally required.
- As is the case with other kinds of evidence, evidence of mechanisms is not infallible.

**Why should one scrutinise evidence of mechanisms in healthcare?** As explained in Sect. 2.3 below, evidence of mechanisms can support or undermine judgements of efficacy and external validity. Therefore, using evidence of mechanisms in concert with other forms of evidence results in better healthcare decisions. (We use the analogy of reinforced concrete to explain this claim; see p. 92.) If this sort of mechanistic reasoning is not properly scrutinised, medical decisions may be adversely affected. For example, current tools for evaluating the quality of clinical research (such as GRADE) do not scrutinise assumptions about mechanisms that have been used to design clinical studies. Just as EBM improved clinical practice by scrutinising clinical studies, scrutinising evidence of mechanisms can lead to further improvements. We have provided some suitable tools for assisting such scrutiny in Chap. 4.

**Mechanisms are the + in EBM+**

## 2.2 What is a Mechanism?

Mechanisms are invoked to explain (Machamer et al. 2000; Gillies 2017b). Textbooks in the biomedical and social sciences are replete with diagrams and descriptions of mechanisms. These are used to explain the proper function of features of the human body, to explain diseases and their spread, to explain the functioning of medical devices, and to explain social aspects of health interventions, among other things.

One kind of mechanism, a *complex-systems mechanism*, is a complex arrangement of entities and activities, organised in such a way as to be regularly or predictably responsible for the phenomenon to be explained (Illari and Williamson 2012). In such mechanisms, spatio-temporal and hierarchical organisation tend to play a crucial explanatory role (Williamson 2018, Sect. 1).

Another kind of mechanism, a *mechanistic process*, consists in a spatio-temporal pathway along which certain features are propagated from the starting point to the end point (Salmon 1998). Examples include the motion of a billiard ball from cue to collision, and the trajectory of a molecule in the bloodstream from injection to metabolism. This sort of mechanism is often one-off, rather than operating in a regular and repeatable way. In the case of environmental causes of disease, the repercussions of these processes may take a long time to develop—e.g., they may be mediated by epigenetic changes.

In the health sciences, mechanistic explanations often involve a combination of these two sorts of mechanism. For example, an explanation of a certain cancer may appeal to the mechanistic processes that bring environmental factors into the human body, the eventual failure of the body's complex-systems mechanisms for preventing damage, and the resulting mechanistic processes that lead to disease, including the propagation of tumours (Russo and Williamson 2012).

We shall use 'mechanism' to refer to a complex-systems mechanism or a mechanistic process or some combination of the two. We should emphasise that mechanisms in medicine and public health may be social as well as biological (see Chap. 9 and Clarke and Russo 2017), and, in the case of medical devices, for instance, they may also include technological components.

A clinical study is the usual method for establishing that two variables are correlated:

A **clinical study** for the claim that  $A$  is a cause of  $B$  repeatedly measures the values of a set of measured variables that includes the variables  $A$  and  $B$ . These values are recorded in a **dataset**. In an **experimental study**, the measurements are made after an experimental intervention. If no intervention is performed, the study is an **observational study**: a **cohort study** follows a group of people over time; a **case control study** divides the study population into those who have a disease and those who do not and surveys each cohort; a **case series** is a study that tracks patients who received a similar treatment or exposure. An **n-of-1 study** consists of repeated measurements of a single individual; other studies measure several individuals. Clinical studies are crucial for estimating any correlation between  $A$  and  $B$ , and they indirectly provide evidence relevant to the claim that  $A$  is a cause of  $B$  (see Fig. 3.1).

On the other hand, a much wider variety of methods can provide good evidence of mechanisms—including direct manipulation (e.g., in vitro experiments), direct observation (e.g., biomedical imaging, autopsy), clinical studies (e.g., RCTs, cohort studies, case control studies, case series), confirmed theory (e.g., established immunological theory), analogy (e.g., animal experiments) and simulation (e.g., agent-based models) (Clarke et al. 2014; Williamson 2018). A mechanistic study is a study which provides evidence of the details of a mechanism:

A **mechanistic study** for the claim that  $A$  is a cause of  $B$  is a study which provides evidence of features of the mechanism by which  $A$  is hypothesised to cause  $B$ . Mechanistic studies can be produced by means of in vitro experiments, biomedical imaging, autopsy, established theory, animal experiments and simulations, for instance. Moreover, consider a clinical study for the claim that  $A$  is a cause of  $C$ , where  $C$  is an intermediate variable on the path from  $A$  to  $B$ —e.g., a surrogate outcome. Such a study is also a mechanistic study because it provides evidence of certain details of the mechanism from  $A$  to  $B$ . A clinical study for the claim that  $A$  is a cause of  $B$  is not normally a mechanistic study for the claim that  $A$  is a cause of  $B$  because, although it can provide indirect evidence that there exists some mechanism linking  $A$  and  $B$ , it does not normally provide evidence of the structure or features of that mechanism. Similarly, a mechanistic study for the claim that  $A$  is a cause of  $B$  is not normally a clinical study for the claim that  $A$  is a cause of  $B$ , because it does not repeatedly measure values of  $A$  and  $B$  together. A study will be called a **mixed study** if it is both a clinical study and a mechanistic study—i.e., if it both measures values of  $A$  and  $B$  together and provides evidence of features of the mechanism linking  $A$  and  $B$ . To avoid confusion, the terminology **clinical study** and **mechanistic study** will be used to refer only to non-mixed studies.

## 2.3 Why Consider Evidence of Mechanisms?

There are various reasons for taking evidence of mechanisms into account when assessing claims in medicine. In general, when evidence is limited, the more evidence one can take into account, and the more varied this evidence is, the more reliable the resulting assessments (Claveau 2013). Moreover, when deciding whether to approve a new health intervention, or whether a chemical is carcinogenic, for example, it can take a very long time to gather enough evidence if the only evidence one considers is clinical study evidence. By considering evidence of mechanisms in conjunction with clinical study evidence, decisions can be made earlier: one can reduce the time taken for a drug to reach market (Gibbs 2000), and reduce the time taken to restrict exposure to carcinogens, for instance.

There are also reasons for considering evidence of mechanisms that are particular to the task at hand. While evidence of mechanisms can inform a variety of tasks (see below), in this book we focus on its use for evaluating efficacy and external validity. Williamson (2018) provides a detailed justification of the need for evidence of mechanisms when performing these two tasks. Here we shall briefly sketch the main considerations.

**Evaluating efficacy.** As noted above, establishing effectiveness can be broken down into two steps: establishing efficacy and establishing external validity. Establishing efficacy, i.e., that *A* is a cause of *B* in the study population, in turn requires establishing two things. First, *A* and *B* need to be appropriately correlated. Second, this correlation needs to be attributable to *A* causing *B*, rather than some other explanation, such as bias, confounding or some connection other than a causal connection (Williamson 2018, Sect. 1).

If it is genuinely the case that *A* is a cause of *B*, then there is some combination of mechanisms that explains instances of *B* by invoking instances of *A* and that can account for the magnitude of the observed correlation. As a mechanism of action may only be present in some individuals but not others, it needs to be credible that the mechanism of action operates in enough individuals to explain the size of the observed correlation in the study population. Just finding a mechanism of action in some individuals is insufficient. Thus, in order to establish efficacy one needs to establish both the existence of an appropriate correlation in the study population and the existence of an appropriate mechanism that can explain that correlation. We shall refer to this latter claim—that there is a mechanism that can explain that correlation—as the general mechanistic claim for efficacy:

**General mechanistic claim.** In the case of efficacy, the general mechanistic claim takes the form: there exists a mechanism linking the putative cause *A* to the putative effect *B*, which explains instances of *B* in terms of instances of *A* and which can account for the observed correlation between *A* and *B*. In the case of external validity, the general mechanistic claim is: the mechanism responsible for *B* in the target populations is sufficiently similar to that responsible for *B* in the study population.

More generally, evidence of mechanisms can help rule in or out various explanations of a correlation. For example, it can help to determine the direction of causation, which variables are potential confounders, whether a treatment regime is likely to lead to performance bias, and whether measured variables are likely to exhibit temporal trends.

Some alternative explanations of a correlation can be rendered less credible by choosing a particular study design. Adjusting for known confounders and randomisation can lower the probability of confounding. Blinding can reduce the probability of performance and detection bias. Larger trials can reduce the probability of chance correlations. Selecting variables *A* and *B* that do not exhibit significant temporal trends and that are spatio-temporally disjoint can reduce the probability of some other explanations.

In certain cases, clinical studies alone might establish that an observed correlation is causal (Williamson 2018, Sect. 2.1). However, establishing a causal claim in the absence of evidence of the details of the underlying mechanisms requires several independent studies of sufficient size and quality of design and implementation which consistently exhibit a sufficiently large correlation (aka ‘effect size’), so as to rule out explanations of the correlation other than causation. This situation is rare: evidence from clinical studies is typically more equivocal. Therefore, evidence of mechanisms obtained from sources other than clinical studies can play a crucial role in deciding efficacy. Considering this other evidence is likely to lead to more reliable causal conclusions. Where this evidence needs to be considered, its quality should be evaluated in ways such as those set out in this book.

**Evaluating external validity.** Having established efficacy, i.e., that a causal relationship obtains in the study population, one needs to establish external validity—that the causal relationship can be extrapolated to the target population of interest.

As noted above, establishing that *A* is a cause of *B* requires establishing both that *A* and *B* are correlated and that there is some mechanism that can account for this correlation. Having established these facts in the study population, one can infer causation in the target population with some confidence if one can establish that:

1. In the target population, there is a mechanism that is sufficiently similar to the mechanism of action in the study population, and
2. Any mechanisms in the target population which counteract this mechanism do not mask the effect of the mechanism of action to such an extent that a net correlation in the target population could not be explained mechanistically.

Evaluating external validity, then, requires evaluating whether the complex of relevant mechanisms in the target population is *sufficiently similar* to that in the study population, in the sense of (1) and (2) holding. Evidence of mechanisms is therefore crucial to this mode of inference.

This form of inference can be especially challenging when the study population is an animal study and the target population is a human population (Wilde and Parkkinen 2017). This is because, despite important similarities between several physiological mechanisms in certain animals and those in humans, many differences also exist. This form of inference can also be challenging when both the study and the target population are human populations. This is because human behaviour is often a component of an intervention mechanism and may in fact hinder the effectiveness of the intervention. We discuss this in Chap. 9. Some well-known examples of behaviour modifying effectiveness include the Tamil Nadu Integrated Nutrition Project (India) and the North Karelia Project (Finland), both discussed by Clarke et al. (2014).

**Other questions.** Apart from when evaluating efficacy and external validity, evidence of mechanisms can also be helpful when:

- Drawing inferences about a single individual, for treatment and personalised medicine (Wallmann and Williamson 2017);
- Commissioning new research and devising new research funding proposals;
- Justifying the use of clinical studies, designing them, and interpreting their results (Clarke et al. 2014);
- Suggesting and analysing adverse drug effects—see Gillies (2017a), who argues that consideration of evidence of mechanisms would have been necessary to avoid the thalidomide disaster;
- Designing drugs and new devices;
- Building economic models in order to ascertain cost effectiveness of a health intervention;
- Deciding how surrogate outcomes are related to outcomes of interest.

**Example. How evidence of mechanisms can help with the analysis of adverse drug effects: abacavir hypersensitivity syndrome.**

Abacavir is a nucleoside analog reverse transcriptase inhibitor, widely used as part of combination antiretroviral therapy for HIV/AIDS, that received an FDA licence in 1998. However, its use was initially complicated by a severe, life-threatening, hypersensitivity reaction that occurred in approximately 5% of users (precise estimates vary; Clay (2002) gives a range of 2.3–9%). There was confusion regarding the cause of this reaction, and it was thought that ‘it is not possible to characterize those patients most likely to develop the HSR’ on the basis of reports of the syndrome (Clay 2002, 2505).

This changed with the discovery that the hypersensitivity syndrome only occurred in individuals with the HLA-B\*5701 allele (Mallal et al. 2002). This discovery arose from evidence of mechanisms. These authors noted that there were similarities between the mechanisms of several hypersensitivity syndromes—by ‘evidence that the pathogenesis of several similar multisystem drug hypersensitivity reactions involves MHC-restricted presentation of drug or drug metabolites, with direct binding of these non-peptide antigens to MHC molecules or haptentation to endogenous proteins before T-cell presentation’ (Mallal et al. 2002, 727). Patients are now genetically screened for the HLA-B\*5701 allele, and this has greatly reduced the incidence of the hypersensitivity syndrome (Rauch et al. 2006).

In this book, we focus largely on the use of evidence of mechanisms to help establish efficacy and external validity. The problem of drawing inferences about a single individual is briefly discussed in Chap. 10.

**Importance of considering evidence of mechanisms.** Recall that in certain cases clinical studies on their own suffice to establish efficacy and there is no need for a detailed evaluation of other evidence of mechanisms. In other cases, however, evidence of mechanisms arising from sources other than clinical studies can be decisive. In such cases, it is important to scrutinise and evaluate this evidence, just as it is important to scrutinise and evaluate clinical studies.

Situations in which it is particularly important to critically assess evidence of mechanisms arising from sources other than clinical studies include:

- Where clinical studies give conflicting results, are of limited quality, or otherwise exhibit uncertainty about the effect size;
- Where randomised clinical studies are not possible, for practical or ethical reasons, in the population of interest (e.g., evaluating putative environmental causes of cancer in humans; evaluating the action of drugs in children and pregnant women);
- Where clinical studies are underpowered with respect to the outcomes of interest (e.g., when assessing adverse reactions to drugs by means of studies designed to test the efficacy of the drug);

- Any question of external validity where clinical studies in the target population are limited or inconclusive;
- Assessing the effectiveness of a public health action or a social care intervention, where a thorough understanding of the relevant social mechanisms is important;
- Assessing the effectiveness of a medical device, where the mechanism of the device and its interaction with biological mechanisms may not be immediately obvious.

Some commentators have argued that one should disregard evidence of mechanisms, largely on the grounds that mechanistic reasoning has sometimes proved dangerous in the past. An infamous example concerns advice on baby sleeping position in order to prevent sudden infant death syndrome (Evans 2002, 13–14). On the basis of seemingly plausible mechanistic considerations, it was recommended that babies be put to sleep on their fronts, since putting a baby to sleep on its back seemed to increase the likelihood of sudden infant death caused by choking on vomit. However, comparative clinical studies later made clear that this advice had led to tens of thousands of avoidable cot deaths (Gilbert et al. 2005). There are several other examples of harmful or ineffective interventions recommended on the basis of mechanistic reasoning (Howick 2011, 154–157). As a result, it has been argued that relying on evidence of mechanisms can do more harm than good.

In many of these cases, however, the proposed evidence of mechanisms was not explicitly evaluated: often, there was little more than a psychologically compelling story about a mechanism (Clarke et al. 2014, 350). In such cases, making the evidence explicit and explicitly evaluating that evidence would have been enormously beneficial. Thus there is a difference between *mechanistic reasoning*, which in some cases is based on rather little evidence and can be problematic, and evaluating *mechanistic evidence*, which is almost always helpful. The case of anti-arrhythmic drugs may help to illustrate this distinction. Arguably, anti-arrhythmic drugs were recommended on the basis of ill-founded mechanistic reasoning (Howick 2011). The story goes as follows. After a heart attack, patients are at a higher risk of sudden death. Those patients are also more likely to experience arrhythmia. On the basis of some mechanistic reasoning, it was thought likely that there was a mechanism linking arrhythmia to heart attacks. Anti-arrhythmic drugs were, as a result, prescribed in an attempt to indirectly prevent heart attacks by directly preventing arrhythmia. It was later discovered on the basis of the Cardiac Arrhythmia Suppression Trial (CAST) that, unfortunately, the drugs led to an increase in mortality (Echt et al. 1991). See also Furberg (1983). However, at least in retrospect, it looks as though insufficient attention had been paid to mechanistic evidence. In particular, there was little reason to think that reducing arrhythmia was a good surrogate outcome for reducing mortality due to heart attacks. Indeed Holman (2017) argues that pharmaceutical company influence was largely responsible for that choice of surrogate outcome. In this case, properly considering the mechanistic evidence may have led to not recommending anti-arrhythmic drugs.

A critic of the use of evidence of mechanisms might respond that even when there exists good evidence of mechanisms, many biomedical processes are so complex that it remains difficult to establish causal claims on the basis of evidence

of mechanisms (Howick 2011, 136–143). For example, there was arguably some good mechanistic evidence in favour of the claim that dalcetrapib lowers the risk of developing coronary heart disease by increasing the ratio of HDL:LDL. However, a randomised controlled trial showed that risk of coronary heart disease was not significantly affected (Schwartz et al. 2012). A possible explanation for this failure was identified by Tardif et al. (2015), who identified two genetic subgroups of patients. While one subgroup appeared to benefit from dalcetrapib, the second genetic subgroup was harmed. Here, while further work was required to understand the mechanisms in play at the stage of the dalcetrapib clinical trial, it appears as if a credible conclusion has now been reached.

More generally, it is widely accepted that the complexity of biomedical processes presents a significant hurdle for establishing causal claims solely on the basis of evidence of mechanisms. But this is exactly why this book recommends explicitly evaluating evidence of mechanisms *alongside* evidence of correlation. Evidence of mechanisms is not sufficient for good clinical decision making—but neither is evidence of mere correlation.

## References

- Clarke, B., & Russo, F. (2017). Mechanisms and biomedicine. In S. Glennan & P. Illari (Eds.), *The Routledge handbook of mechanisms and mechanical philosophy*. Chap. 24 (pp. 319–331). London: Routledge.
- Clarke, B., Gillies, D., Illari, P., Russo, F., & Williamson, J. (2014). Mechanisms and the evidence hierarchy. *Topoi*, 33(2), 339–360.
- Claveau, F. (2013). The independence condition in the variety-of-evidence thesis. *Philosophy of Science*, 80(1), 94–118.
- Clay, P. (2002). The abacavir hypersensitivity reaction: A review. *Clinical Therapeutics*, 24(10), 1502–14.
- Echt, D. S., Liebson, P. R., Mitchell, L. B., Peters, R. W., Obias-Manno, D., & Barker, A. H., et al. (1991). Mortality and morbidity in patients receiving encainide, flecainide, or placebo. *New England Journal of Medicine*, 324(12), 781–788.
- Evans, D. (2002). Database searches for qualitative research. *Journal of the Medical Library Association*, 90, 290–3.
- Furberg, C. (1983). Effect of antiarrhythmic drugs on mortality after myocardial infarction. *The American Journal of Cardiology*, 52, C32–C36.
- Gibbs, J. B. (2000). Mechanism-based target identification and drug discovery in cancer research. *Science*, 287(5460), 1969–1973.
- Gilbert, R., Salanti, G., Harden, M., & See, S. (2005). Infant sleeping position and the sudden infant death syndrome: systematic review of observational studies and historical review of recommendations from 1940 to 2002. *International Journal of Epidemiology*, 34(4), 874–887.
- Gillies, D. (2017a). Evidence of mechanism in the evaluation of streptomycin and thalidomide. *Studies in the History and Philosophy of Biological and Biomedical Sciences*, 66, 55–62.
- Gillies, D. (2017b). Mechanisms in medicine. *Axiomathes*, 27(6), 621–634.
- Holman, B. (2017). Philosophers on drugs. *Synthese*. <https://doi.org/10.1007/s11229-017-1642-2>.
- Howick, J. (2011). *The philosophy of evidence-based medicine*. Chichester: Wiley-Blackwell.
- Illari, P. M., & Williamson, J. (2012). What is a mechanism? Thinking about mechanisms across the sciences. *European Journal for Philosophy of Science*, 2, 119–135.

- Machamer, P., Darden, L., & Craver, C. (2000). Thinking about mechanisms. *Philosophy of Science*, 67, 1–25.
- Mallal, S., Nolan, D., Witt, C., Masel, G., Martin, A., Moore, C., et al. (2002). Association between presence of HLA-B\*5701, HLA-DR7, and HLA-DQ3 and hypersensitivity to HIV-1 reverse-transcriptase inhibitor abacavir. *The Lancet*, 359(9308), 727–732.
- Pawson, R. (2003). Nothing as practical as a good theory. *Evaluation*, 9(4), 471–490.
- Rauch, A., Nolan, D., Martin, A., McKinnon, E., Almeida, C., & Mallal, S. (2006). Prospective genetic screening decreases the incidence of abacavir hypersensitivity reactions in the Western Australian HIV cohort study. *Clinical Infectious Diseases*, 43(1), 99–102.
- Russo, F., & Williamson, J. (2012). EnviroGenomarkers: The interplay between mechanisms and difference making in establishing causal claims. *Medicine Studies: International Journal for the History, Philosophy and Ethics of Medicine and Allied Sciences*, 3, 249–262.
- Salmon, W. C. (1998). *Causality and explanation*. Oxford: Oxford University Press.
- Schwartz, G. G., Olsson, A. G., Abt, M., Ballantyne, C. M., Barter, P. J., Brumm, J., et al. (2012). Effects of dalcetrapib in patients with a recent acute coronary syndrome. *New England Journal of Medicine*, 367(22), 2089–2099.
- Tardif, J.-C., Rhéaume, E., Perreault, L.-P. L., Grégoire, J. C., Zada, Y. F., Asselin, G., Provost, S., Barhdadi, A., Rhains, D., L'Allier, P. L., Ibrahim, R., Upmanyu, R., Niesor, E. J., Benghozi, R., Suchankova, G., Laghrissi-Thode, F., Guertin, M.-C., Olsson, A. G., Mongrain, I., Schwartz, G. G., & Dubé, M.-P. (2015). Pharmacogenomic determinants of the cardiovascular effects of dalcetrapib. *Circulation: Genomic and Precision Medicine*, 8(2), 372–382.
- Wallmann, C., & Williamson, J. (2017). Four approaches to the reference class problem. In G. Hofer-Szabo & L. Wroński., editor, *Making it formally explicit: Probability, Causality, and Determinism*. Dordrecht: Springer.
- Wilde, M., & Parkkinen, V.-P. (2017). Extrapolation and the Russo–Williamson thesis. *Synthese*. <https://doi.org/10.1007/s11229-017-1573-y>.
- Williamson, J. (2018, in press). Establishing causal claims in medicine. *International Studies in the Philosophy of Science*. <http://blogs.kent.ac.uk/jonw/>.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



# Chapter 3

## How to Consider Evidence of Mechanisms: An Overview



**Abstract** This chapter introduces how to assess evidence of mechanisms, explaining a summary protocol for use of evidence of mechanisms in assessing efficacy, then external validity (developed theoretically in Part III, with tools for implementation offered in Part II). An outline of quality assessment—of a whole body of evidence, rather than individual studies—is given. The chapter finishes with a brief introduction to the ideas developed in Part III: gathering evidence of mechanisms (Chap. 5); evaluating evidence of mechanisms (Chap. 6); and using evidence of mechanisms to evaluate causal claims (Chap. 7).

This section summarises the overall approach taken in this book. It develops some of the more practical issues raised in the introduction (Chap. 1) and begins to attach these to the more theoretical discussions found in Part III. We start with an overview of the way in which effectiveness can be evaluated. As discussed above, effectiveness can be evaluated by evaluating efficacy and external validity. A translation of the core ideas of this chapter to other arenas of practice, such as social policy, is readily possible—although we do not attempt this here in the interest of clarity.

### 3.1 Questions to Address

The following protocol can be used to test a causal claim:

**Efficacy**

*Does the effect size and quality of clinical studies establish that the observed correlation is causal?*

Yes? Efficacy is established.

No?

- Evaluate other evidence for the claim that there exists an appropriate mechanism that can explain the observed correlation.
  - What are the hypothesised mechanisms?
  - How well confirmed is each such mechanism? What are the gaps? How well confirmed is each feature (process, entity, activity and organisational feature) of the mechanism?
  - Can the mechanism account for the full effect size? Are there counteracting mechanisms? What is the evidence that the influence of any counteracting mechanisms is less than that of the proposed mechanism?
- Evaluate other evidence to rule in or out other explanations of the correlation. Are any remaining explanations better confirmed than the hypothesis that the correlation is causal?

*Efficacy is established if one can establish, in the study population, the existence of a correlation and the existence of a mechanism that can explain this correlation.*

**External validity**

*Do clinical studies directly establish a suitable association and mechanism in the target population?*

Yes? Effectiveness is established.

No?

- Evaluate the claim that the mechanism of action is sufficiently similar in the target and study populations.
- Evaluate the claim that in the target population, any counteracting mechanisms that are not also present in the study population do not mask the effect of the mechanism of action.
- Evaluate other evidence for a correlation in the target population.

*External validity is established if one can establish similarity of relevant mechanisms in the study and target populations, and thereby establish, in the target population, the existence of a correlation and the existence of a mechanism that can explain this correlation.*

In the case of efficacy, it is rare that clinical studies *alone* establish that the observed correlation is causal in the study population. Clinical research does not (generally) take place in isolation from basic science research. Many aspects of the design and interpretation of clinical trials—such as the choice of outcome measures, therapeutic regimes compared, and patient recruitment criteria—are influenced by evidence of mechanisms. Thus, even in the absence of complete knowledge of the underlying mechanisms, evidence of mechanisms contributes to establishing efficacy (Illari 2018). This is also true with respect to external validity, where it is almost never the case that clinical studies in the study population will directly establish both a suitable association and mechanism that will apply to the target population. Rather, external validity inferences proceed in one of the following ways (Parkkinen and Williamson 2017):

1. By identifying and comparing key details of the mechanisms in the study and target populations.
2. Inductively, by observing a similar effect in many different experimental populations and generalizing from these to the target population (Wilde and Parkkinen 2017).
3. Phylogenetically, by identifying the mechanism in the study population, and then inferring that the mechanisms in the study and target population are similar due to shared ancestry of the populations. The greater the degree of isolation between the target population and the study population, the less reliable this inference will be.
4. By means of a robustness argument: showing that the mechanism in the study population is so robust, and differences between the study and target populations are so minor, that the mechanism of action will also obtain in the target population.

Thus, for both efficacy and external validity one typically needs to consider evidence of mechanisms arising from sources other than the clinical studies that establish a correlation in the study population. This means that those who evaluate evidence will generally need to consider [mechanistic studies](#), in addition to clinical studies, in order to make causal judgements.

Of course, some features of a putative mechanism may already be well established, in which case there will usually be no need to revisit the evidence for those features. Other features will be more contentious. It is only by explicitly identifying these features and the evidence that pertains to them that one can critically appraise a proposed mechanism.

## 3.2 Quality of Evidence and Status of Claim

**Quality of evidence.** Evidence for various claims can be ranked by quality. We distinguish three main kinds of claim: claims about correlation, claims about mechanisms and causal claims (including claims about efficacy and claims about external validity). We use the scale in Table 3.1 to rank the quality of this evidence.

**Table 3.1** Quality levels of evidence, based on Atkins et al. (2004)

Quality level	Interpretation
High	Further research is highly unlikely to have a significant impact on our confidence in the claim
Moderate	Further research is moderately unlikely to have a significant impact on our confidence in the claim
Low	Further research is moderately likely to have a significant impact on our confidence in the claim
Very low	Further research is highly likely to have a significant impact on our confidence in the claim

Note that this ranking system evaluates the *total body of evidence* pertaining to the claim in question. This is in sharp contrast to other EBM methods that evaluate single studies in isolation.

This approach to ranking quality on the basis of stability of confidence can be found in the original GRADE framework (see Guyatt et al. 2008). According to this sort of approach, establishing a causal claim requires confidence in the *stability* of that causal claim, in addition to confidence in the nature of the claim itself. We should emphasise that the interpretation of each category concerns the *in principle possibility* of obtaining further research that changes confidence in the claim. A brief example will be helpful here. Suppose current evidence warrants 75% confidence in a causal claim. One then learns that there is further evidence which warrants a 25% change in confidence, but one does not know the direction of this change. i.e., one does not know whether this new evidence warrants 50% confidence or 100% confidence. The 75% confidence is not sufficiently stable for the claim to be considered established or even provisionally established. This is because future evidence may be likely to decide between the 50 and 100% confidence, leading to a large change in confidence either way.

GRADE later changed their interpretation of quality levels, dropping reference to the likelihood that further evidence will change confidence in the claim (Balshem et al. 2011, Table 2; Hultcrantz et al. 2017). This was because of concerns about the situation in which further evidence is unlikely to be obtained in practice: if further research is unlikely to be carried out then further research is unlikely to have an impact on our confidence in the causal claim in question. This change is unnecessary: as noted above, the key question is whether evidence can *in principle* be obtained to significantly alter confidence in the claim. In short, just because ethical or practical considerations make it very unlikely that further research on a particular claim will be carried out, that does not imply that current evidence is high quality.

**Table 3.2** Status of a claim

Status	Interpretation
Established	A claim is established when community standards are met for adding the claim to the body of evidence—i.e., for granting the claim and treating it as evidence for other claims In order to establish a claim, evidence must warrant a high level of confidence in the claim and this evidence must itself be high quality
Provisionally established/ provisional	Moderate quality evidence warrants a high level of confidence in the claim
Arguably true/ arguable	The claim is neither established nor provisionally established, but evidence of at least moderate quality warrants significantly more confidence in the claim than in its negation, or low quality evidence warrants a high level of confidence in the claim
Speculative	A claim is speculative if it falls into none of the other categories
Arguably false	The claim is neither ruled out nor provisionally ruled out, but evidence of at least moderate quality warrants significantly more confidence in the negation of the claim than in claim itself, or low quality evidence warrants a high level of confidence in the negation of the claim
Provisionally ruled out	Moderate quality evidence warrants a high level of confidence in the negation of the claim
Ruled out	A claim is ruled out when community standards are met for adding the negation of the claim to the body of evidence In order to rule out a claim, high quality evidence must warrant a high level of confidence in the negation of the claim

**Status of claim.** In addition to the quality of the evidence, we shall also be concerned with the status that the evidence confers on the claim under consideration. The status of a claim will be measured on the scale depicted in Table 3.2.

Note that this table invokes two separate levels: the quality level applies to the total evidence, while the level of confidence applies to the claim in question. The status of the claim depends on both the quality of the evidence as well as the degree of confidence that the evidence warrants.

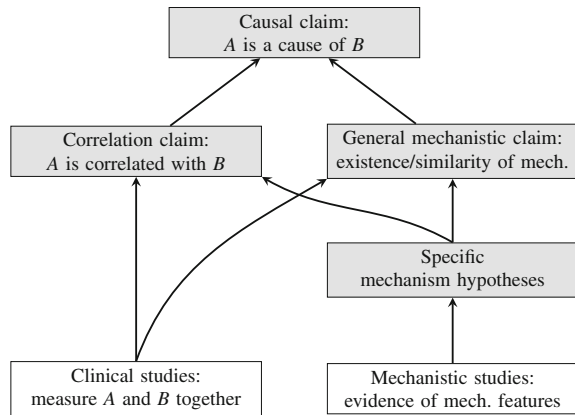
We will see shortly that the status of a causal claim will depend on the status of a correlation claim (assessed, e.g., by using the GRADE system) together with the status of a mechanism claim (assessed by the procedures outlined in Chap. 6).

Appendix B provides a simple probabilistic interpretation of the notion of quality and status developed in this section.

### 3.3 Overall Approach

Figure 3.1 depicts the evidential relationships linking the concepts of this book; cf. Williamson (2018b). A claim that *A* is a cause of *B* is assessed by evaluating

**Fig. 3.1** The evidential relationships employed in this book. See Williamson (2018b)



two further claims. The first—the **correlation** claim—is the claim that  $A$  and  $B$  are appropriately correlated. The second is the **general mechanistic claim**. In the case of efficacy, this is the claim that there exists an appropriate mechanism linking  $A$  and  $B$  that can explain  $B$  in terms of  $A$  and that can account for the extent of the correlation. There are two ways of confirming this general mechanistic claim: either via clinical studies which find a correlation that can only be explained by the general mechanistic claim being true, or by identifying key features of the actual mechanism of action, which are confirmed by mechanistic studies. In the case of external validity, the general mechanistic claim is the claim that the mechanisms of action in the study and target population are sufficiently similar. Again, this can be confirmed either by clinical studies on both populations that find similar correlations, or by ascertaining key features of the mechanism of action in each population and finding that these are similar. In addition, clinical studies provide good evidence of correlation, and, in certain circumstances, an established mechanism of action can also provide good evidence of correlation (Williamson 2018a, Sect. 2.2).

There is a **correlation** between two variables  $A$  and  $B$  if these two variables are probabilistically dependent, i.e.,  $P(B|A) \neq P(B)$ . In many situations where a causal relationship is being assessed, the correlation claim of interest is the probabilistic dependence of  $A$  and  $B$  conditional on some set of a priori potential confounding variables. A confounding variable is a variable correlated with both  $A$  and  $B$ , such as a common cause of  $A$  and  $B$ . Note that ‘correlation’ is sometimes used to refer to a linear dependence; here we use the term in the more general sense to refer to any probabilistic dependence.

**Specific mechanism hypothesis.** This is a hypothesis of the form: a specific mechanism with features  $F$  links the putative cause to the putative effect.

In contrast, other current EBM methods for evidence appraisal focus almost exclusively on the evaluation of clinical studies, i.e., on the two arrows at the bot-

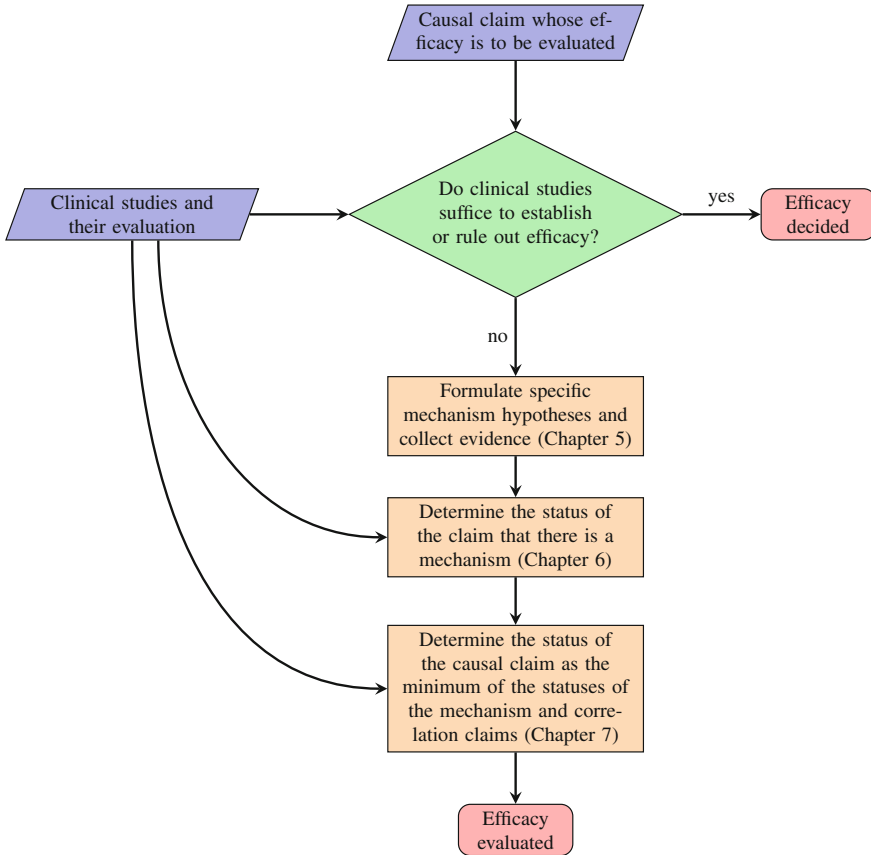


Fig. 3.2 Evaluating efficacy

tom left of Fig. 3.1. Moreover, they tend to conflate these two arrows—they do not distinguish the role of clinical studies in evaluating a correlation claim from their role in determining whether there is some underlying mechanism of action. Once these two roles are separated, it is clear that mechanistic studies also need to be appraised when evaluating the latter general mechanistic claim. This is the **evidential pluralism** introduced in Sect. 1.1.

Two flowcharts summarise the overall approach. Figure 3.2 depicts the workflow when evaluating efficacy. The second flowchart, Fig. 3.3, applies to the evaluation of external validity. In each case there are three principal steps: gathering evidence of mechanisms; evaluating evidence of mechanisms; and using evidence of mechanisms to evaluate causal claims. Procedures for implementing the three steps are developed in Chaps. 5, 6 and 7 respectively. The main ideas can be summarised as follows.

**Gathering evidence of mechanisms** (Chap. 5). It is typically more difficult to find evidence of mechanisms in the literature than it is to find relevant evidence of cor-

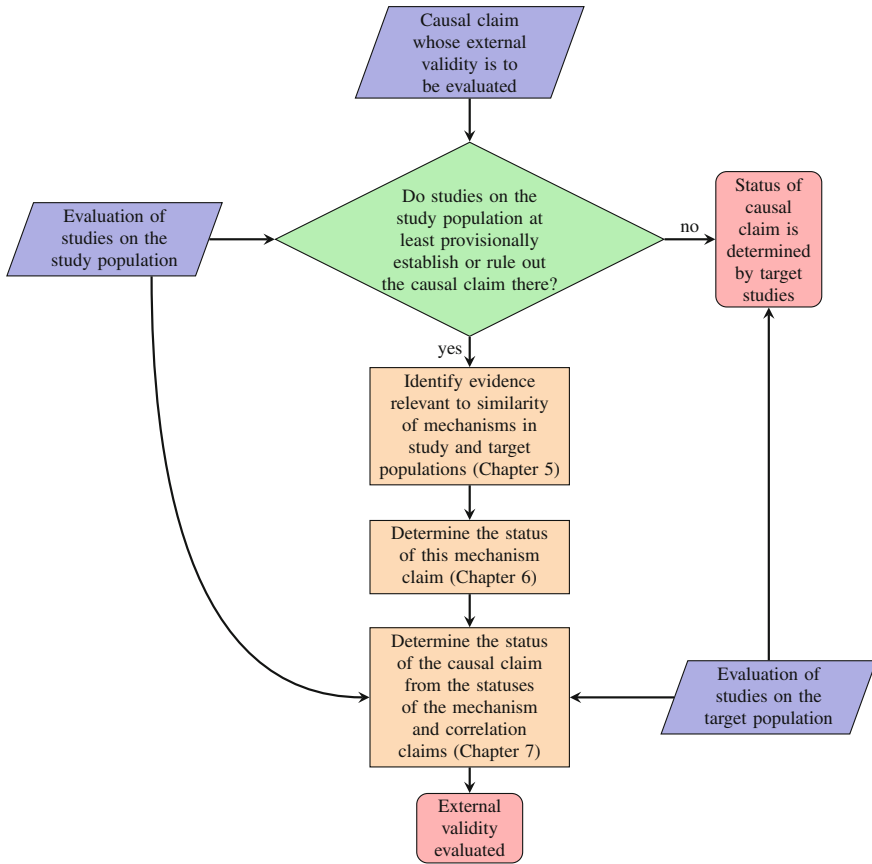


Fig. 3.3 Evaluating external validity

relation. This is because evidence of mechanisms is characteristically produced by mechanistic studies, and there are a large number of diverse types of mechanistic study (Smith et al. 2016). This makes the process of recognising good evidence more difficult, because an investigator is likely to be unfamiliar with the details of all the possible kinds of research that might be relevant to a clinical outcome. Historically, as Evans (2002) has argued, database indexing practices for these studies have tended to be unsystematic in comparison with those for clinical studies. Arguably, this has contributed to a tendency to overlook or entirely ignore evidence of mechanisms that arises from sources other than clinical studies.

However, as explained above, such evidence of mechanisms is often crucial to establishing efficacy and external validity. Given this, the difficulties in gathering evidence of mechanisms need to be overcome. As a first step towards overcoming the difficulties, we propose a five-step strategy for identifying evidence of mechanisms, a strategy that in part relies upon existing evidence of mechanisms:

1. *Identify*: Identify a number of specific mechanism hypotheses.
2. *Formulate*: For each specific mechanism hypothesis, formulate a number of review questions.
3. *Search*: Use these review questions to search the literature.
4. *Refine*: Identify the evidence most relevant to the mechanism hypothesis in question by refining the results of this search.
5. *Present*: Present the evidence relevant to the mechanism hypothesis.

This strategy is intended to help overcome some of the practical difficulties of identifying evidence of mechanisms—difficulties which may prevent practitioners from considering all the evidence. We develop this strategy in more detail in Chap. 5. We have also provided a series of tools in Part II that help users conduct certain parts of this process in specified areas of practice.

**Evaluating evidence of mechanisms** (Chap. 6). In evaluating the quality of mechanistic evidence, the following questions are likely to be most helpful.

1. *How well established and understood are the methods by which the evidence (of existence of a mechanism or some of its features) was produced?*

Well established methods whose functioning and potential biases are properly understood and which can be calibrated against other well established methods typically provide higher quality evidence than methods that rely on novel techniques that cannot be calibrated against better understood methods.

2. *Can the item of evidence be produced by independent methods?*

Employing several detection techniques and checking their results against each other is a common way to distinguish experimental artefacts from valid results. (The greater the number of independent methods that can confirm a result, the higher the quality of an item of evidence.)

3. *Are the model systems that are used in experimental research well characterised?*

Model systems do not usually exactly reproduce the relevant human mechanisms. Have the relevant differences been characterised for the system(s) used in this research?

4. *Can the mechanism be observed operating in many different background contexts?*

The more robust a mechanism is against variation in background conditions, the less likely it is that inferences based on evidence of the mechanism will err because of unknown contextual factors interfering with the mechanism. Demonstrable robustness of the mechanism itself thus makes for higher quality evidence.

Sections 6.1 and 6.2 describe a procedure for evaluating the quality of mechanistic studies that is broken down to three steps:

1. Evaluating methods
2. Evaluating the implementation of methods
3. Evaluating results

The **status** of the general mechanistic claim is then assigned as follows. A mechanism to account for efficacy can be considered *established* in two ways. First, when high quality clinical studies exhibit a substantial correlation that is not explainable by, e.g., confounding or bias. Alternatively, when there are high quality mechanistic studies that confirm all the crucial component features of the mechanism. A hypothesised mechanism for efficacy is considered *ruled out* when there is high quality evidence against the existence of the component features of the mechanism. A mechanism may also be ruled out if high quality clinical studies consistently fail to show results one would expect if the mechanism was operating as hypothesised. A mechanism to account for external validity is considered *established* when high quality evidence establishes the similarity of all the crucial components of the mechanism in the study and target populations. A mechanism hypothesised to account for external validity is considered *ruled out* when there is high quality evidence of dissimilarity of mechanisms between the study and target populations. The more gaps or inconsistencies there are in the evidence base for a particular claim about a mechanism, the lower its status.

There are other useful status indicators that require slightly more careful judgement. *Provisionally established* claims admit some gaps in the evidence base, but require overall a good amount of high quality evidence. *Arguable* claims have evidence in their support that is either moderate quality or that has important gaps. *Speculative* claims are supported by evidence that shows mixed results, or have little evidence in their support beyond theoretical intuition or speculation.

These issues are explained in more detail in Chap. 6.

**Using evidence of mechanisms to evaluate causal claims** (Chap. 7). Having ascertained the status of a correlation claim and relevant mechanism claims, one can use these to determine the status of the causal claim of interest. This process, which is explored in Chap. 7, may be summarised as follows.

In order to establish efficacy, one needs to establish that the putative cause and effect are correlated and that there is a mechanism that can account for this correlation. More generally, one can take the status of a causal claim to be the minimum of the status of the correlation claim and the status of the general mechanistic claim. For instance, if a correlation is arguable but the existence of any underlying mechanism is provisionally ruled out, then the causal claim itself is provisionally ruled out.

Turning to external validity, the situation is more complicated because one needs to consider (i) evidence for the causal claim obtained directly on the target population, (ii) evidence for efficacy in the study population, and (iii) evidence of similarity of mechanisms between the study and the target populations. Evidence directly about the target may be boosted (or undermined) by observing that efficacy does (or does not) hold in a study population that shares similar mechanisms with the target population. Table 7.1 combines the status of the causal claim in the target with the status of efficacy in the study and the status of the claim that the mechanisms in the target and the study are similar.

## References

- Atkins, D., Best, D., Briss, P. A., Eccles, M., Falck-Ytter, Y., Flottorp, S., et al. (2004). Grading quality of evidence and strength of recommendations. *BMJ*, 328(7454), 1490–1490.
- Balshem, H., Helfand, M., Schünemann, H. J., Oxman, A. D., Kunz, R., Brozek, J., et al. (2011). Grade guidelines: 3. Rating the quality of evidence. *Journal of Clinical Epidemiology*, 64(4), 401–406.
- Evans, D. (2002). Database searches for qualitative research. *Journal of the Medical Library Association*, 90, 290–293.
- Guyatt, G. H., Oxman, A. D., Vist, G. E., Kunz, R., Falck-Ytter, Y., Alonso-Coello, P., et al. (2008). GRADE: An emerging consensus on rating quality of evidence and strength of recommendations. *British Medical Journal*, 336(7650), 924–926.
- Hultcrantz, M., Rind, D., Akl, E. A., Treweek, S., Mustafa, R. A., Iorio, A., et al. (2017). The grade working group clarifies the construct of certainty of evidence. *Journal of Clinical Epidemiology*, 87(Supplement C):4–13.
- Illari, P. M. (2018). Who's afraid of mechanisms? *British Journal for Philosophy of Science*.
- Parkkinen, V.-P., & Williamson, J. (2017). Extrapolating from model organisms in pharmacology. In A. La Caze & B. Osimani (Eds.), *Uncertainty in pharmacology: Epistemology, methods, and decisions*. Dordrecht: Springer.
- Smith, M. T., Guyton, K. Z., Gibbons, C. F., Fritz, J. M., Portier, C. J., Rusyn, I., et al. (2016). Key characteristics of carcinogens as a basis for organizing data on mechanisms of carcinogenesis. *Environmental Health Perspectives*, 124, 713–721.
- Wilde, M., & Parkkinen, V.-P. (2017). Extrapolation and the Russo–Williamson thesis. *Synthese*. <https://doi.org/10.1007/s11229-017-1573-y>.
- Williamson, J. (2018a). Establishing causal claims in medicine. *International Studies in the Philosophy of Science*, in press. Preprint available at <http://blogs.kent.ac.uk/jonw/>.
- Williamson, J. (2018b). Establishing the teratogenicity of Zika and evaluating causal criteria. *Synthese*, in press. Preprint available at <http://blogs.kent.ac.uk/jonw/>.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



**Part II**  
**Tools for Working with Mechanisms**

# Chapter 4

## Tools



**Abstract** If theoretical developments in evidence assessment are to prove useful, guidance on implementation is essential, and this chapter fills that need. A variety of tools are offered, which can be used either in isolation, or in the various combinations suggested. The starting point is an *Is your policy really evidence-based?* tool which should be very widely usable to give a very quick overview. Then two tools are offered for guideline developers for medical practice; these offer improved assessment of evidence of mechanism in assessing clinical trials, and, if needed, in basic science papers. For politicians, journalists, academics, and so on, a critical appraisal tool is offered alongside GRADE-style tables for mechanism assessment. A final tool is designed specifically for public health and social care.

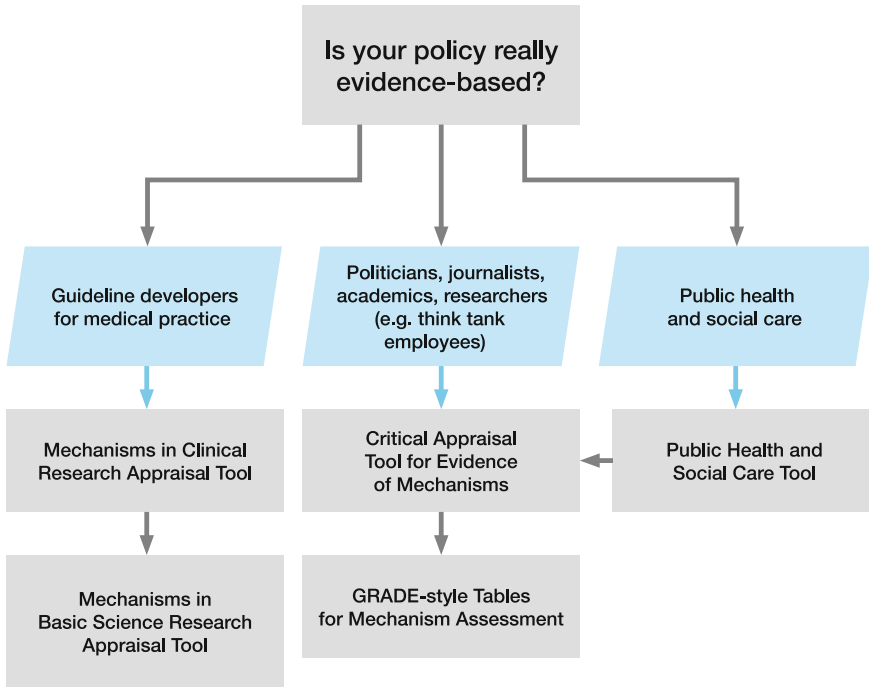
In this chapter, we present a number of tools for evaluating evidence of mechanisms that have been tailored for different users. A flowchart that shows how these tools can be used together is presented below in Fig. 4.1.

### 4.1 Introduction

#### How to use these tools

For most users, the *Is your policy really evidence-based?* tool (Sect. 4.2) will be the best place to start, because it can give a quick indication of cases where a more detailed review of evidence might be valuable. If a policy is found to have possible weaknesses in its underlying evidence base, the user can then employ the other tools provided here to produce a more thorough account of the strengths and weaknesses of the policy's evidence base. While we encourage interested users to experiment with each of these tools to see which might best fit their purposes, we propose the following provisional plan:

**For those interested in guidelines for medical practice.** We would encourage these users to move on to a more systematic review of evidence using the *Mechanisms in Clinical Research appraisal tool* (see Sect. 4.3). This might also involve a more



**Fig. 4.1** A suggested work-flow for using the tools presented here

detailed review of evidence arising from basic science work using the *Mechanisms in Basic Science Research appraisal tool* (see Sect. 4.4).

**For those working on public health and social care guidelines.** The *Public Health and Social Care* tool (in Sect. 4.7) would be the most natural place to begin, because it explicitly asks appraisers to evaluate evidence of mechanisms that pertains both to individuals and to groups. Because of the diversity of the underlying research in public health and social care policies, the *Critical Appraisal Tool for Evidence of Mechanisms* would be the most useful tool to apply (see Sect. 4.5).

**For those interested in other policies, such as politicians, journalists, and academics.** The most natural way to proceed would depend on the nature of the policy in question. If the policy is largely medical (i.e. dealing with the effects of an intervention on an individual, with a largely biological theme) then the *Mechanisms in Clinical Research appraisal tool* would be appropriate (see Sect. 4.3), perhaps followed by the *Mechanisms in Basic Science Research appraisal tool* (see Sect. 4.4). Otherwise, the *Critical Appraisal Tool for Evidence of Mechanisms* (Sect. 4.5) could be used in combination with the *GRADE-style Tables for Mechanism Assessment* (Sect. 4.6) as a next step.

### **Limitations of these tools**

These tools are fallible, and their use is not a substitute for expert appraisal of a guideline or policy. Answering each of the steps requires user judgement, and the scores produced by each tool contribute to—rather than determine—the quality of recommendations. In other words, the tool alone will not provide a final and complete judgement of the quality of evidence, and their use is not a substitute for expert judgement.

These tools are specifically designed to assist in the evaluation of causal relationships. Guidance that relies on the [precautionary principle](#) may therefore score poorly, just because the precautionary principle is used when evidence of causal relationships is limited. Those poor scores should not therefore be interpreted as sufficient to alter such guidance.

These tools are currently beta versions that are suitable for testing. They have been tested by the EBM+ team during development. We welcome feedback on these tools via the EBM+ website at [ebmplus.org](http://ebmplus.org). Feedback will help inform the next version of these tools, which will be accessible from the EBM+ website.

## **4.2 Is Your Policy Really Evidence-Based?**

### **Introduction**

This is a tool for appraising a wide range of policy decisions. Policies are likely to be more effective when they are based on evidence. But there are many kinds of evidence, and many ways to use evidence. Just as not all kinds of evidence are created equal, not all ways of using evidence are equally good. This tool permits the user to draw rapid but useful conclusions about the evidence that a particular policy is based on, and the way that it is based on this evidence.

Policies that use different kinds of evidence together, in an explicit and careful way, are generally better justified than policies that do not. This tool allows the user to quickly and fairly judge whether their policy is evidence-based in this way. Whilst the effectiveness of a policy is somewhat dependent on the strength of its evidence, other factors are also significant. These include proper implementation, strict adherence, and the responsiveness of policy updates.

### **Who should use this tool**

This tool is a light-touch and rapid means of appraising the way that a recommendation is supported by its evidence. It is intended for use on existing policies, rather than being a tool for those constructing recommendations in the first instance. The tool was written largely with medicine and social care in mind. For example, it asks questions about evidence from basic science research because this plays an important role in supporting policy in those areas. However, we acknowledge that other types

of information are used in concert with evidence from scientific research in building policies—this tool can accommodate a wide range of different needs and different stakeholder groups working with issues of medical policy. It is envisaged that civil servants, activists, political parties, What Works Centres, and guideline developers will find this tool useful. The tool in table 4.1 might also be valuable in other areas (such as evaluating economic policy) with appropriate translation.

To provide some examples of ways that this tool might be used:

- Clinicians in primary care, public health, or social care, might use this tool as a check when considering the implementation of a new clinical guideline, or in other situations where rapid appraisals of guidelines might be otherwise helpful (for example, in multidisciplinary team meetings).
- Patient groups might use this tool to aid discussions of new treatment recommendations.
- Journalists might use the tool to begin investigating controversial policy decisions.
- Guideline authors might use this tool as a first step when considering revisions to existing guidance.
- Decision makers in local authorities might choose to use this tool when making decisions about service provision.
- Politicians (and their teams) could use this tool to evaluate their manifesto claims—or those of their opponents.
- Directors of social care and public health might use this tool to evaluate existing practices.
- This tool would be useful as part of a post-hoc effectiveness evaluation tool-kit that could be applied to policies in the event of their failure.

It is important to remember that policy evolves and develops out of many actions and involves many actors. This is true in democratic societies (where these interactions are usually at least partially visible). It is also true in more closed societies, where it is less easy to observe. In both cases, evidence and its appraisal are but one part of the mix. The relationships have been studied by political scientists (Kingdon and Thurber 1984), by policy makers themselves (National Audit Office 2003) as well as social scientists more generally (Nutley et al. 2000). There are plentiful models describing the process (Cooksey 2006; Ogilvie et al. 2009). The relationship between evidence and policy is a complex one. This has to be acknowledged, but that notwithstanding, it is important to apply the highest standards of evaluation we can to the available evidence.

### **How to use this tool**

This tool should be used when examining a specific policy or recommendation. For example, we might be interested in examining a claim that, for disease  $x$  in population  $y$  use drug  $z$ . This policy will (hopefully) be supported by some group of research evidence that shows that drug  $z$  is the most effective treatment for disease  $x$ .

**Table 4.1** Is your policy really evidence-based tool

<p><b>Question 1: Is the policy or recommendation explicit?</b></p> <p>A recommendation can't be evidence-based unless it is explicit. For instance, the recommendation 'be more healthy' cannot be evidence-based unless it specifically describes how you should be healthy.</p> <p>Yes ? Somewhat ? No ?</p>	<p><b>Question 4a: Is the recommendation based on suitable evidence of correlations?</b></p> <p>Suitable evidence here might include clinical trials or observational studies.</p> <p>Yes ? Unclear or limited ? No ?</p>	<p><b>Question 6: Are the limitations of the evidence explicit in the recommendation?</b></p> <p>The explicit expression of the limitations of the evidence is usually linked to more consistent recommendations (eg, noting patterns of co-morbidity in trial subjects in guidelines).</p> <p>Yes ? Unclear or partial expression of limitations ? No ?</p>
<p><b>Question 2: Is the recommendation based on empirical studies of some kind?</b></p> <p>Evidence is empirical. Unless a recommendation is based on empirical observations it cannot be evidence based. Simulations, models, and well-verified theories count as empirical evidence here.</p> <p>Yes (named or described studies) ? Somewhat (studies not named or described) ? No ?</p>	<p><b>Question 4b: Is the recommendation based on suitable evidence of mechanisms?</b></p> <p>Are laboratory studies or basic science research used as part of the evidence-base for the recommendation?</p> <p>Yes ? Unclear or limited ? No ?</p>	<p><b>Question 7: Does the policy or recommendation offer comparisons to other recommendations?</b></p> <p>New recommendations are either intended to replace older recommendations or give guidance that is more appropriate for a particular population (geographical, demographic...). Is the recommendation compared to others in either of these ways?</p> <p>Yes (or not applicable) ? Unclear or partial discussion ? No ?</p>
<p><b>Question 3: Is the policy or recommendation made for a suitable population?</b></p> <p>Different populations may have different needs. A recommendation is supported by evidence if it was produced in a group of research participants known to be similar to the target population.</p> <p>Yes (similar populations) ? Maybe (populations not described or known) ? No (populations known to be dissimilar) ?</p>	<p><b>Question 5: Are the sources of evidence integrated together in the recommendation?</b></p> <p>Integrating sources of evidence leads to stronger recommendations. For example, has the selection of biomarkers in a trial been integrated with current best laboratory evidence?</p> <p>Yes ? Unclear or limited ? No ?</p>	<p><b>Overall Score</b></p> <ul style="list-style-type: none"> <li>■ A red box in Questions 1 or 2 suggests that the recommendation is poorly evidence-based.</li> <li>■ A red box (or 3 yellow boxes) by Questions 3, 4 or 5 suggests that the recommendation is somewhat evidence-based.</li> <li>■ No red boxes (or less than 3 yellow boxes) by Questions 6 or 7 suggest that the recommendation is strongly evidence-based.</li> </ul>

The tool then asks users a series of questions that reveal difficulties in the evidential support for that policy. These are ranked because failures in the early questions reveal more serious difficulties than failures in the later questions. These steps correspond to aspects of the account of how to gather, evaluate, and use, evidence of mechanisms that is developed in Part III. There are seven steps, each with a simple traffic-light checklist (green, yellow, or red) and each of which reflects one aspect of the relationship between the recommendation and the evidence base. The overall score for a particular policy can then be expressed by recording the lowest numbered step in which the red box is checked. For example, a policy would score 3 if it were found to be based on research on a population that was extremely unlike the intended population for its use. Note that if no red boxes are checked for any of the questions then the overall score should be noted as 7+, indicating that a policy is as evidence-based as possible. Multiple yellow flags should indicate caution, and we suggest that when three or more yellow flags are present, the score should be recorded as equal to the stage at which the third yellow flag is indicated. This overall score gives an extremely concise measure of the strength of the links between the evidence-base and the recommendation. A fuller appraisal of the policy can also be easily seen by consulting the full page of scores for each step. These initial appraisals can then form a basis for more detailed appraisal using other tools, as detailed in Sect. 4.1.

### 4.3 Mechanisms in Clinical Research Appraisal Tool

#### Introduction

This tool presents a method that a researcher would use to evaluate a group of clinical research publications. The aim of this method is to facilitate the construction of concise summaries of the mechanistic aspects of a group of clinical research publications. These summaries can then be used by a panel of experts in the context of making policy decisions about healthcare in combination with other data extraction tools (such as GRADE). Note that this tool is not intended to produce a full reconstruction of all the mechanisms that might be relevant. Instead, the summaries are intended to reveal the mechanistic aspects of clinical research. For example, some understanding of the hypothesised mechanism of action of a drug will inform the design of a clinical trial testing that drug. These mechanistic assumptions should be considered when interpreting this clinical trial.

This tool is comparatively simple, and therefore is intended for use in circumstances where the details of a mechanism are thought likely to be straightforward. In cases where either a) the consequences of a policy decision are rather serious (such as making decisions about medicines for use in pregnancy) or b) when the research base that grounds a body of clinical research is disputed or complex (such as the evaluation of treatments for chronic fatigue syndrome) we suggest that a more detailed appraisal be conducted using our *Mechanisms in Basic Science Research appraisal tool* (see Sect. 4.4). A more theoretical approach to the appraisal process can be found in Fig. 5.1.

### Who should use this tool

This tool is intended for use during the development of clinical guidelines. Parts of this tool can be used by different groups as the process of guideline development proceeds. The data extraction parts of this tool (A1–A3, as well as A5 if used) may be of use to literature review specialists alongside existing appraisal work. While these parts of the tool do assume some expertise in dealing with the medical literature, we do not assume domain-specific expertise in these parts. Parts of this tool—particularly A4—do assume a higher level of expertise in some specified scientific domain, and this stage will largely be carried out by domain experts. Finally, A6 is intended to be carried out by those with expertise in producing guidance from clinical research.

This tool has been designed with the current (2018) practices of NICE as an archetype. We understand that practices vary in different contexts, and that the demands of a different context of practice might produce difficulties in using this tool.

### How to use this tool

We describe a six-stage method for using this tool. The numbers of the stages (e.g. A3) are also shown on the flowchart in Fig. 4.2 to assist in understanding the overall appraisal process. Each of the stages will help evaluate the evidence-base that supports (or undermines) a drug's safety and efficacy. Note that not all steps will be necessary in each case. Instead, this process is adaptable to suit cases where the evidence base is favourable, or cases where the evidence base is unfavourable, or cases where the evidence base is more mixed. Note too that different stages of the process are likely to be carried out by different evaluators. We have designed this tool to assist smooth transitions between evaluators. An overview of the intended process is below.

**A1: collate clinical studies** At this stage, the process is identical to that of traditional publication screening. A set of search terms should be selected, and applied to published and unpublished studies. Duplicates should be excluded, and then appropriate selection criteria (e.g. study language, age of study) should be applied. This will result in a group of clinical studies that we call the *appraisal stack*.

**A2: extract data relating to mechanisms from these studies** Using Table 4.2, data should then be extracted from this stack of clinical studies. This will serve to identify both the content and quality of these studies. Again, we envisage that this step will accompany existing data collection protocols that are used in guideline development. Data collection should take place for each one of the reviewed studies, and a data summary table containing data summaries for each article should be produced.

**A3: review data for gaps** Using the completed data summary table, the analyst can then make some preliminary recommendations regarding the set of clinical research papers as a whole. These tools will particularly help to determine whether there are problems about the mechanistic aspects of this corpus of literature. We foresee several different possibilities at this stage that might require different handling.

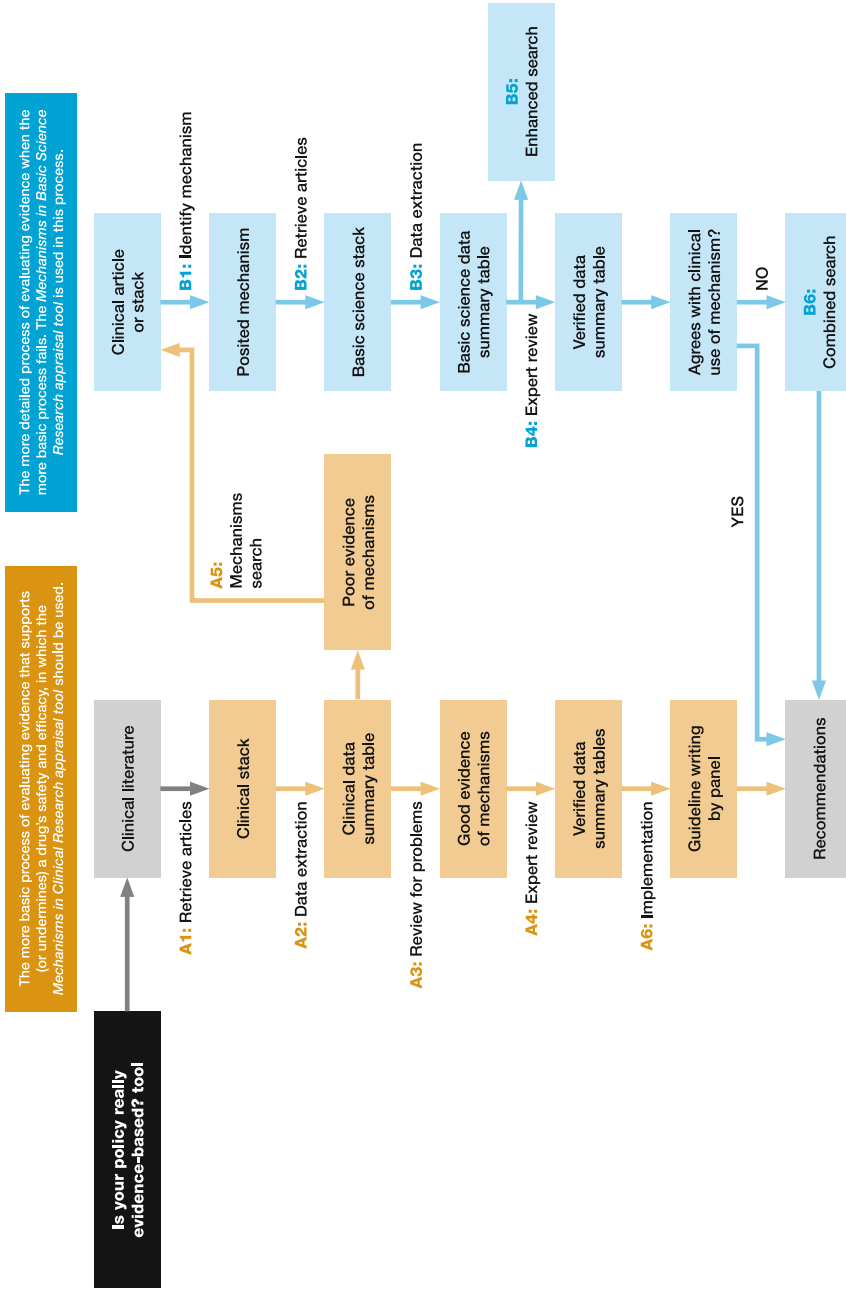


Fig. 4.2 A suggested work-flow for the integrated use of the clinical and basic science tools

**Table 4.2** Clinical research data extraction table

Study Details	Published?	Study Design	Mechanism Hypothesised?	Intervention used	Outcome measure	Comparison	Population
Author(s), title, summary	<input type="checkbox"/> Yes – journal	<input type="checkbox"/> Cluster RCT	<input type="checkbox"/> No mechanism posited	Describe intervention(s) used:	Describe the outcome measure used:	Describe the comparison(s) performed:	Describe the trial population:
	<input type="checkbox"/> Yes – other	<input type="checkbox"/> Individual RCT	<input type="checkbox"/> Mechanism posited: no evidence provided (please provide details)				
	<input type="checkbox"/> No	<input type="checkbox"/> Correlation study					
	<input type="checkbox"/> Unclear	<input type="checkbox"/> Cross sectional study					
DOI:	<input type="checkbox"/> Prospective cohort study <input type="checkbox"/> Retrospective cohort study	<input type="checkbox"/> Mechanism posited: evidence cited (please provide citations)	Is the choice of intervention explicitly justified? <input type="checkbox"/> No <input type="checkbox"/> Yes (please provide details)	Is the choice of outcome measure explicitly justified? <input type="checkbox"/> No <input type="checkbox"/> Yes (please provide details)	Is the comparison clinically relevant (eg. appropriate doses of comparator drugs)? <input type="checkbox"/> Yes (please provide details) <input type="checkbox"/> No (please provide details)	Does this differ from the intended treatment population? <input type="checkbox"/> Yes (please provide details) <input type="checkbox"/> No (please provide details)	
Date:	<input type="checkbox"/> Other, please describe						

**Established mechanisms:** in cases where a group of clinical research papers appears to be explicitly based on a known mechanism, and where there is ample discussion of that mechanism in the basic science research literature, no further investigation will generally be required, and the user should proceed directly to stage four. A special case might be where the clinical studies appear to rely on the same mechanism, but where there is no explicit justification of that mechanism. Users in this case should make explicit note of this, and refer the issue to an expert panel (**A4**) as a possible precursor to a more developed mechanism search.

**Other cases:** in cases where the clinical research literature does not link neatly to an established mechanism, a more detailed search for a mechanism will generally be helpful to guideline authors. In this case, proceed to **A5**.

**A4: expert review** The data summary table should now be passed to domain experts for review. One important question at this stage is to ensure that the selection of publications examined at stage **A3** is fair and unbiased. So the experts should satisfy themselves that no cherry-picking of the research literature has taken place, and that the data extraction has fairly summarised the state of knowledge in the relevant field. If this is not the case, proceed to **A5** to conduct a more detailed mechanism search. If the domain experts are satisfied, this verified data summary table can then be passed on to a guidelines panel for use in their deliberations in **A6**.

**A5: mechanism search** Conduct a more detailed mechanism search using the Mechanisms in Basic Science Research Appraisal Tool to address gaps in the clinical research literature. This will frequently require consultation with domain experts for search term scoping and expert review. Once complete, the mechanisms data, together with the clinical data summary table, should be passed to an expert review panel for approval before moving to **A6**.

**A6: implementation/recommendation/review stage** The data summary table should then be used, in concert with other data extraction tools (and, if applicable, a summary of mechanisms data), in formulating recommendations. Here, the data summary tool is designed to facilitate panel discussions about the strengths and weaknesses of individual studies, as well as to assist with more overarching decisions about recommendations.

As discussed above, use of the *Mechanisms in Basic Science Research* tool may be necessary in some appraisals. Figure 4.2 provides an overview of the integrated use of these two tools.

## 4.4 Mechanisms in Basic Science Research Appraisal Tool

### Introduction

This tool presents a method that a researcher would use to evaluate a mechanistic claim about a drug treatment as it appears in the basic science literature. The aim is to facilitate the construction of concise summaries for a group of basic science publications. These summaries can then be used alongside similar summaries of clinical research by a panel of experts in the context of making policy decisions. Note that this tool is not intended to produce a full reconstruction of all the mechanisms that might be relevant. Instead, the summaries will indicate the degree to which the published evidence supports some mechanism. As mechanisms frequently inform the design and interpretation of clinical trials, these summaries of evidential support for mechanistic claims that might be found in clinical research will enable a policy panel—with appropriate expert input—to appropriately evaluate both clinical and basic science research together in an integrated way.

This tool is comparatively detailed, and is therefore largely intended for use in circumstances where the details of a mechanism are particularly contentious. Broadly, this might be when either a) the consequences of a policy decision are rather serious (such as making decisions about medicines for use in pregnancy) or b) when the research base that grounds a body of clinical research is disputed or complex (such as the evaluation of treatments for chronic fatigue syndrome). Mechanisms of interest in more simple cases are likely to be dealt with adequately by our *Mechanisms in Clinical Research appraisal tool*.

### Who should use this tool

This tool is intended for use during the development of clinical guidelines. Parts of this tool can be used by different groups as the process of guideline development proceeds. The data extraction parts of this tool (**B2** and **B3**) are likely to be largely carried out by literature review specialists alongside existing appraisal work. While these parts of the tool do assume some expertise in dealing with the medical literature, we do not assume domain-specific expertise in these parts. Parts of this tool, particularly **B1**, **B4**, and **B6**, do assume a higher level of expertise in some specified scientific domain, and this stage will largely be carried out by domain experts. Finally, **B1** and **B6** will generally require close collaboration between literature review specialists, and domain experts.

### How to use this tool

We describe a six-stage method for using this tool. Not all steps will be necessary in each case. We generally intend this tool to follow on from issues identified during the use of the *Mechanisms in Clinical Research Appraisal Tool* (see Sect. 4.3), and this guide assumes that this is the case. Please also see the overview flowchart (Fig. 4.2) to understand the overall appraisal process.

**B1: identify a posited mechanism** Begin with a clinical research paper (or appraisal stack from the clinical tool). Then retrieve citations from the clinical paper(s) that describe key assumptions about mechanisms. These might include:

- Mechanism of action
- Biomarkers
- Patient population recruitment criteria
- Surrogate outcome measures

If no mechanism is described in the clinical research paper(s), or if the user is using this tool independently of the clinical research tool, expert advice is desirable at this stage to assist with the identification of a mechanism.

**B2: retrieve papers** Retrieve basic science papers (identified in **B1**). Then identify the purpose that these basic science papers are used for in the relevant clinical paper(s).

**B3: data extraction** Using Table 4.3, extract data from the relevant basic science papers identified in **B2**. Repeat for all basic science papers.

**B4: expert review** Pass data tables to experts for review to verify that the extraction has fairly summarised the relevant field. One important question at this stage is to ensure that the selection of publications examined at stage **B3** is fair and unbiased. Domain experts should satisfy themselves that no cherry-picking of the research literature has taken place. If extraction has not fairly summarised the field then proceed to **B5**. If however the experts are satisfied, then this verified data can then be passed to the guidelines panel for use in their deliberations. If problems and inconsistencies are revealed during this process, proceed to **B6**.

**B5: enhanced search (for cases where the cited literature is unrepresentative of a field)** Conduct a keyword search on the mechanism (see also Chapter 5). This should then be followed by applying stages **B1** to **B4** to the updated group of basic science papers found by this keyword search.

**B6: combined search (for cases where the clinical and basic sciences literature are divergent)** Conduct a combined search across both clinical and basic science material, concentrating on the connection between different kinds of evidence with respect to a claim. This will require input from experts for both the clinical and basic science material.

Once completed, the data summaries from this tool should be passed back to the relevant guideline panel, ideally in combination with the relevant clinical data summary table.

Table 4.3 Basic science data extraction tool

Description	Score/Reported detail		Possible red flags
	Basic Information		
Short citation or DOI	http://dx.doi.org		No DOI, predatory journal, not peer reviewed
Retracted or corrected	<input type="checkbox"/> Not known to be retracted <input type="checkbox"/> Retracted <input type="checkbox"/> Corrected (please describe)		Retracted publications, substantial correction or correspondence
Age of publication	X years		Physically outdated research, eg. papers that precede major theoretical or technological changes
Does the conclusion support the citation?	<input type="checkbox"/> Main conclusion supports citation <input type="checkbox"/> Conclusion does not support citation		Conclusion does not support citation
Was this conclusion an intended finding of the research? Is the citation an incidental feature of the research paper? Does the conclusion arise from subgroup or post hoc analysis?	<input type="checkbox"/> Citation is an incidental feature of the study/ (please describe below)		Citation was an incidental finding of the study, raising issues of powering etc.
Might there be a mismatch between the basic science research and its use in clinical research (eg. use of biomarkers, outcome measures)	Please describe		
<b>Types of experiment</b>			
What is the research design?	<input type="checkbox"/> Experimental <input type="checkbox"/> Observational <input type="checkbox"/> Whole animal study <input type="checkbox"/> Molecular study <input type="checkbox"/> Other (please describe)	<input type="checkbox"/> Other/unclear <input type="checkbox"/> Simulation <input type="checkbox"/> Cell-based	Design-target mismatch or design-conclusions mismatch
Are the limitations of the research explicitly discussed and evaluated?	<input type="checkbox"/> Evaluated <input type="checkbox"/> Mentioned <input type="checkbox"/> Absent		
Controlled experiment?	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unclear		
Type of experimental probe?	Please describe:		Actually or accurately described
New or established experimental method?	<input type="checkbox"/> Established <input type="checkbox"/> Obsolete <input type="checkbox"/> New (please describe)	<input type="checkbox"/> Unclear	Cautions for either potentially obsolete methods, or highly novel methods without discussion of calibration and validation
<b>Mechanism posited</b>			
Is a mechanism described or posited in the paper? If so, what is it?	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unclear Brief description if possible		
Does the experimental finding reported bear on some aspect of that mechanism?	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unclear		Experimental findings appear unrelated to discussed mechanism
Are links drawn to relevant literature to support the mechanism described in the paper?	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unclear		Striking absence of relevant work
<b>Relevance to humans</b>			
Does the paper make comments about the relevance of the research to humans?	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unclear		Human case known to be different (eg. beta-lactam antibiotic research in rodents)
Is that extrapolation explicitly justified?	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unclear		

## 4.5 Critical Appraisal Tool for Evidence of Mechanisms

### Introduction

This tool presents a method for critical appraisal of mechanistic evidence which is modelled on the EBM critical appraisal worksheets publicly available at the [Oxford Centre for Evidence-Based Medicine website](#). This aim is to provide a integrated way of evaluating the processes of gathering, evaluating, and using, evidence of mechanisms to determine the status of a causal claim. The tool is intended to be used in a stand-alone way, ideally in concert with an evaluation of other forms of evidence that might bear on a causal claim of interest. The theoretical details of these evaluations are explained in later parts of this book (see Chaps. 5, 6, and 7 respectively).

### Who should use this tool

The tool is fairly, rather than very, detailed. It is a sensible next-step from the *Is Your Policy Really Evidence-Based tool* (Sect. 4.2) for many purposes, although we would particularly recommend it as a tool for use in contexts that are not directly related to developing healthcare guidelines. The *Mechanisms in Clinical Research appraisal tool* (Sect. 4.3) would be better fitted to these purposes (Table 4.4).

### How to use this tool

The tool consists of eight questions. Each is accompanied with a note of guidance about both how to interpret the question (and showing how the specific question fits in with the evaluation process), as well as some notes of guidance about where to find information that will contribute to answering the question posed. Together, these questions can help reveal the strength of evidential support for some specific mechanism hypothesis.

## 4.6 GRADE-Style Tables for Mechanism Assessment

### Introduction

One widely used approach to assessing and summarizing quality of evidence and strength of recommendations in systematic reviews and clinical practice guidelines is the Grading of Recommendations Assessment, Development and Evaluation (GRADE) system (Guyatt et al. 2011), used for example by NICE (NICE 2014). The GRADE process involves collecting evidence to address a specific question about specific outcomes, and rating the quality of evidence according to the quality of study design, risk of bias, imprecision, inconsistency of findings, indirectness (relative to the target population), and magnitude of effect. The quality of evidence and strength of recommendation is then summarized in a table. GRADE tables do not include an

**Table 4.4** A critical appraisal tool for evidence of mechanisms

What is best?	Where do I find the information?
<b>What question is to be addressed with evidence of mechanisms?</b>	
<p>The main question being addressed should be clearly formulated so that one can identify mechanistic studies relevant to the causal claim of interest. The question should clearly indicate (1) the causal claim of interest, (2) whether the question concerns efficacy or external validity, (3) the general mechanistic claim relevant to the question, and (4) one or more specific mechanism hypothesis for which evidence is sought.</p>	<p>The <b>title</b>, <b>abstract</b>, and <b>introduction</b> of a research article typically state the outcome and the proposed mechanism that is considered in a study.</p>
<b>Have all the relevant mechanistic studies been identified?</b>	
<p>The starting point for a comprehensive search for relevant studies is the major bibliographic databases (such as PubMed), but should also include a search of reference lists that appear in relevant articles, and contact with experts, particularly to enquire about unpublished studies. The search should not be limited to the English language literature.</p>	<p>For guidance, refer to <a href="#">Chapter 4 of Evaluating evidence of mechanism in medicine</a>.</p>
<b>What are the criteria used to select articles to be included in the evaluation?</b>	
<p>The inclusion and exclusion criteria should be appropriate for the question being addressed, and should be defined in advance of the search. These criteria should specify whether the studies included have demonstrated that the effect/outcome is responsive to the proposed mechanisms, and in which way they are. For example, the criteria should specify whether the studies included demonstrated dose-dependency or not; whether <i>in vitro</i>, <i>in vivo</i>, or both types of study were included; or whether human, experimental animal, or both types of study were included.</p>	<p>The <b>methods</b> section of an article should describe the inclusion and exclusion criteria. The <b>abstract</b> and <b>results</b> sections should describe the results of the evaluation.</p>
<b>Were the included studies sufficiently valid for the type of question asked?</b>	
<p>One should evaluate the quality of each study by using predetermined quality criteria appropriate for the type of study. Due to the nature of mechanistic studies, these considerations vary case-by-case, but typically include assessment of biological preparation and other methods. One should consider whether the methods used can be taken to be reliable and appropriate regarding the objectives of a study.</p>	<p>The <b>methods</b> section of an article should describe the preparation methods, what was measured and how. Studies that do not transparently describe how the results were obtained should be excluded.</p>
<b>Were the results similar from study to study?</b>	
<p>Ideally, the results of different studies should be similar, but in many cases there are several different mechanism hypotheses relevant to the outcome of interest. If the latter is the case, one should evaluate which of these mechanisms are best supported by the evidence reviewed.</p>	<p>The <b>results</b> section of a review article should state whether the results are different from study to study. Similarly, the results section of individual studies should describe what the results are. When assessing many individual studies, one should compare the results across the studies to evaluate the concordance of the evidence.</p>

(continued)

**Table 4.4** (continued)

What is best?	Where do I find the information?
<b>What is the status of each specific mechanism hypothesis?</b>	
The review of evidence should determine the status of each specific mechanism hypothesis.	Considerations to take into account when evaluating the status of a mechanistic hypothesis can be found in <a href="#">Chapter 6 of <i>Evaluating evidence of mechanisms in medicine</i></a> .
<b>What is the status of the general mechanistic claim?</b>	
Once the status of each specific mechanism hypothesis has been determined, one should combine the evidence of mechanisms with evidence of correlation to arrive at a status assessment for the causal claim of interest.	A procedure for combining evidence from mechanistic studies and evidence from clinical studies to determine the status of the general mechanistic claim is described in <a href="#">Chapter 6 of <i>Evaluating evidence of mechanisms in medicine</i></a> .
<b>What is the status of the causal claim?</b>	
Once the status of the mechanism hypothesis has been determined, one should combine the evidence of mechanisms with evidence of correlation to arrive at a status assessment for the causal claim of interest.	A procedure for combining evidence of mechanisms and correlation to evaluate the status of a causal claim is described in <a href="#">Chapter 7 of <i>Evaluating evidence of mechanisms in medicine</i></a> .

explicit assessment of mechanistic evidence. In this tool we provide some examples of ways in which one might extend GRADE evidence profile tables to also include evidence of mechanisms. The proposed amendments are modelled according to the categories used in the GRADE tables. These amended tables illustrate that it is possible to incorporate many aspects of the approach of this book into a popular system like GRADE, without having to make any radical changes.

**GRADE-style table for mechanism assessment**

**Who should use this tool**

This tool is intended for use in cases where a systematic review of evidence is being conducted as part of policy development. Thus this tool is intended for a fairly expert audience, with the assumption that users will be generally familiar with current best practice in evidence appraisal. This tool is therefore an ideal step-up from the less thorough assessment that a researcher might have produced using either the *Is your policy really evidence-based?* (Sect. 4.2) and/or the *Critical appraisal tool for evidence of mechanisms* (Sect. 4.5).

**How to use this tool**

Table 4.5 provides a template for an augmented GRADE-style table. We assume that a user is generally familiar with the current GRADE method for evidence appraisal. This augmented table is intended to be used a similar way. However, as it contains some questions which are likely to be unfamiliar, we have provided some notes of guidance here on these proposed new categories.

**Table 4.5** Grade table with mechanism assessment

<b>Interventions</b>						
<b>Outcome:</b>						
<b>Assessment: clinical studies</b>						
No. of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	No. of events
						Effect
						Certainty
<b>Outcome 2</b>						
No. of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	No. of events
						Effect
						Certainty
<b>Assessment: mechanism</b>						
<b>Mechanism hypothesis</b>						
			Gaps		Masking	Inconsistency
						Indirectness
						Quality and status
<b>Overall assessment</b>						

Note that providing answers to these questions may require substantial investigation, particularly in cases where the relevant mechanisms are unclear or disputed. The *Clinical Research* (Sect. 4.3) and *Basic Science* (Sect. 4.4) tools may be of value in such cases.

**Mechanism hypothesis.** If the quality of clinical studies is high, and observed effect sizes sufficiently large, there may be no need to formulate and evaluate specific mechanism hypotheses. Otherwise, each specific hypothesised mechanism should be sketched here.

**Gaps.** Crucial features of the specific mechanism hypothesis that are lacking evidence, or for which there is high risk that the available evidence is biased due to methodological limitations of the studies.

**Masking.** Evidence of mechanisms that counteract the effect of the hypothesized mechanism. This will reduce the plausibility of the intervention having a robust effect through the proposed mechanism.

**Inconsistency.** Evidence for feature(s) of a mechanism is inconsistent when there is some evidence in favour of a feature of a mechanism, and some against it, or when there is evidence for two or more mutually exclusive mechanisms. Note that inconsistency should be evaluated taking into account the amount and quality of evidence—e.g., if some of the conflicting evidence is systematically significantly less reliable due to study limitations, the inconsistency is not to be considered as severe.

**Indirectness.** Evidence relating to other populations and evidence of crucial differences between mechanisms in those populations and mechanisms in the target population.

In the **quality and status** box, one should state the overall quality of the mechanistic studies and the status of the specific mechanism hypothesis given the evidence (see Sect. 3.2 and Chap. 6). Any outstanding study limitations can be summarized here.

The **overall assessment** box should include an evaluation of the status of the general mechanistic claim, and should discuss how this informs the overall assessment of the status of the effectiveness claim. See Sect. 6.3 and Chap. 7.

### Worked example

Table 4.6 depicts a worked example of this GRADE-style appraisal, which is an assessment of brief contact interventions for reducing self-harm. Further worked examples can be found in [Appendix C](#).

**Table 4.6** GRADE-style table showing assessment of brief contact interventions for reducing self-harm; c.f. Milner et al. (2016)

Brief contact interventions		Repeated episode of self-harm or suicide attempt		Assessment: clinical studies		12 months or less		No. of events	Effect	Certainty
No. of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	No. of events	Effect	900 / 7585	0.87 (0.74 to 1.04)	@@@⊖Moderate
11	RCT	No serious risk of bias	No serious inconsistency	No serious indirectness	Some serious imprecision					
Assessment: mechanism										
Mechanism hypothesis		Gaps		Masking		Inconsistency		Indirectness		Quality and status
Social support: "BCIs provided participants with a sense of connectedness and the sense they were being listened to"		Serious gap: the link to reduction of harm		No serious masking		No serious inconsistency		No serious indirectness		@@@⊖Moderate, Status: Arguable
Suicide prevention literacy: "BCIs improved an individual's knowledge about suicidal behaviours or self-harm (eg, risk and protective factors), what help is available, and how to access this help"		Serious gap: the link to reduction of harm		No serious masking		No serious inconsistency		No serious indirectness		@@@⊖Moderate, Status: Arguable
Learning alternative behaviours: "BCIs involved participants learning positive and functional alternative behaviours to self-harm"		Serious gaps: limited evidence of the link from the intervention; the link to reduction of harm		No serious masking		No serious inconsistency		No serious indirectness		@@@⊖Low, Status: Speculative
Overall assessment										
<p>Evidence of moderate quality suggests that brief contact interventions reduce the number of repeat episodes of self-harm or suicide attempts; however, the results are not statistically significant and the existence of a correlation can at best be considered arguable. Evidence of high quality shows brief contact interventions improve social support and suicide prevention literacy. These specific links are considered established. Evidence of moderate quality suggests brief contact interventions may also provide people with alternative coping behaviours. This specific link is arguable. However, the links from these intermediate points to the endpoint of interest, reduction of self-harm or suicide attempts, have not been investigated in detail. Overall, therefore, the specific mechanism hypotheses are arguable or speculative. Accordingly, the causal claim is at best arguable. Further research is needed to explore alternative mechanisms that could be targeted for more efficacious interventions, and to explain the lack of impact of brief contact interventions on self-harm and suicide.</p>										

## 4.7 Public Health and Social Care Tool

### Introduction

This is a tool for appraising public health and social care policies, which differ in many ways from the kinds of interventions that are used in clinical medicine. This tool will help the authors and evaluators of these policies ensure that their interventions are as closely connected to underlying research in the relevant sectors (Fig. 4.3) as possible. Users of this tool may find the discussion of mechanisms in public health in Chap. 9 a helpful adjunct to this tool.

### Public Health and Social Care tool

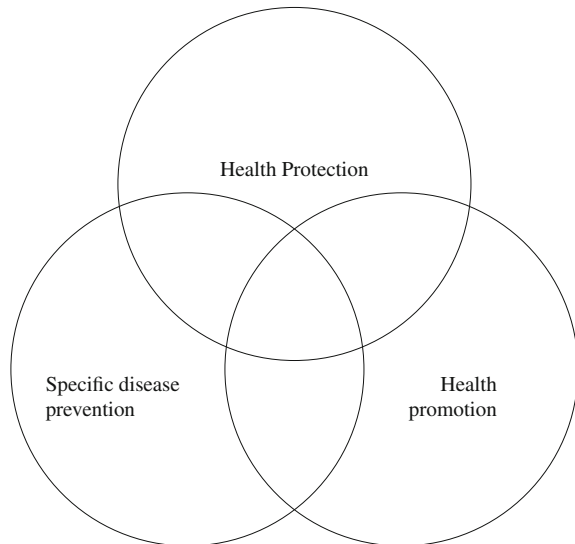
#### Who should use this tool

This tool is largely aimed at experts in public health and social care policy. It assumes a fairly high level of knowledge of the research that might be relevant for appraising a policy, and requires the user to exercise their judgement in evaluating that evidence. It is also a comparatively detailed process. A better alternative tool for contexts where a lighter review of evidence is thought to be sufficient is the *Is your policy really evidence-based?* tool found in Sect. 4.2.

#### How to use this tool

This tool can be employed as a way of checking the alignment between the available evidence of mechanisms and policy guidance. It is thus intended to help resolve problems regarding the external validity of research, and will help researchers be confident that their recommendations will be applicable to their population of interest.

**Fig. 4.3** Our understanding of public health (after Tannahill, 1985)



Note that the tool presupposes that population-based research (such as trials of an intervention) will be evaluated using other methods such as GRADE.

Part one of the tool (Table 4.7) asks the user to provide three sets of preliminary information: about the public health problem that the proposed intervention is intended to affect, about the nature of the intervention itself, and about the population that this intervention is meant to be applied to.

Part two of the tool (Table 4.8) then asks the user to answer questions about the evidence that bears on each of these preliminary information from part one. These questions about the evidence are divided along two axes—individual/group and biological/social. Ideally, the user should be satisfied that there are no identifiable problems in either of the four quadrants.

Note that the questions in the tools may be hard to answer in some cases. For example, research on social mechanisms may be lacking. Or, for new risks, the research base might be very slender. To offer a note of reassurance from our testing, difficulties in gathering relevant research should be regarded as a positive finding in the context of this tool.

Other parts of this book may be a helpful addition to this tool, depending on the case at hand. The *Critical Appraisal Tool for Evidence of Mechanisms* (in Sect. 4.5)

**Table 4.7** Part one: preliminary questions for Public Health and Social Care appraisal

<b>What is the proposed intervention?</b>	<b>or</b>	<b>Which health outcome or behaviour is the focus of this intervention?</b>
<b>What are the desired health outcomes or behaviour changes?</b>	<b>or</b>	<b>What are the proposed interventions?</b>
<b>In what population is the intervention to be used?</b>		

**Table 4.8** Part two: evidence questions for Public Health and Social Care appraisal

Biological questions about groups	Social questions about groups
Which aspects of the biological mechanism(s) are known? Which are not? Why might public health interventions targeting the pathogens fail?	Which socio-economic or psychological mechanisms are known? Which are not? What is the public perception of the disease in terms of risk, seriousness and personal vulnerability?
Which mechanisms come into play as a population or different segments in the population react to an intervention?	
Biological questions about individuals	Social questions about individuals
Are there sub-groups within the population that should be specifically targeted?	How can behavioural mechanisms reduce exposure?
How can they be reached and what specific mechanisms might come into play?	

and the *GRADE-style Tables for Mechanism Assessment* (in Sect. 4.6) would be particularly appropriate next-steps.

## References

Cooksey, D. (2006). *A review of UK health research funding*. London: The Stationery Office.

Guyatt, G., Oxman, A. D., Akl, E. A., Kunz, R., Vist, G., Brozek, J., et al. (2011). GRADE guidelines: 1. Introduction—GRADE evidence profiles and summary of findings tables. *Journal of clinical epidemiology*, *64*(4), 383–394.

Kingdon, J. W., & Thurber, J. A. (1984). *Agendas, alternatives, and public policies*. Boston, MA: Little, Brown.

Milner, A., Spittal, M. J., Kapur, N., Witt, K., Pirkis, J., & Carter, G. (2016). Mechanisms of brief contact interventions in clinical populations: a systematic review. *BMC Psychiatry*, *16*(1), 194.

National Audit Office (2003). *Getting the evidence: Using research in policy making*. Technical report, National Audit Office, London, HMSO.

NICE (2014). Developing NICE guidelines: The manual. Published online. <https://www.nice.org.uk/process/pmg20/chapter/introduction-and-overview>.

- Nutley, S. M., Smith, P. C., & Davies, H. T. (2000). *What works?: Evidence-based policy and practice in public services*. Bristol: Policy Press.
- Ogilvie, D., Craig, P., Griffin, S., Macintyre, S., & Wareham, N. J. (2009). A translational framework for public health research. *BMC Public Health*, 9(1), 116.
- Tannahill, A. (1985). What is health promotion? *Health education journal*, 44(4), 167–168.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



**Part III**  
**Core Principles**

## Chapter 5

# Gathering Evidence of Mechanisms



**Abstract** In this chapter we put forward more theoretical proposals for gathering evidence of mechanisms. Specifically, the chapter covers the identification of a number of mechanism hypotheses, formulation of review questions for search, and then how to refine and present the resulting evidence. Key issues include increased precision concerning the nature of the hypothesis being examined, attention to differences between the study population (or populations) and the target population of the evidence assessors, and being alert for masking mechanisms, which are other mechanisms which may mask the action of the mechanism being assessed. An outline example concerning probiotics and dental caries is given. (Databases that may be helpful for some searches can be found online in [Appendix A](#)).

In the next three chapters, we develop core principles for evaluating efficacy and external validity. In this chapter we put forward proposals for gathering evidence of mechanisms. Then, in Chap. 6 we discuss how to evaluate this evidence. In Chap. 7, we explain how this evaluation can be combined with an evaluation of correlation in order to produce an overall evaluation of a causal claim.

In the case of efficacy, where clinical studies find a correlation between the putative cause and effect, the task is to determine whether this correlation is causal by looking for further evidence of mechanisms. In order to evaluate efficacy, it is necessary to determine the status of the [general mechanistic claim](#), i.e., to ask whether the correlated putative cause and effect are also linked by a mechanism that can account for the extent of the observed correlation.

In the case of external validity, the existing evidence may establish causality in a study population that differs from the target population of interest. Here the relevant [general mechanistic claim](#) that needs to be evaluated is that mechanisms in the study and target population are sufficiently similar.

**General mechanistic claim for efficacy.** In formulating the general mechanistic claim for efficacy, the following questions should be addressed:

- What is the relevant population?
- What is the intervention or exposure level?
- What is the outcome and how is it measured?

**General mechanistic claim for external validity.** In determining the general mechanistic claim concerning external validity, the following questions should be addressed:

- What is the target population? What is the study population?
- What is the intervention or exposure level in the target?
- What is the outcome and how is it measured in the target?
- What is the intervention or exposure level in the study?
- What is the outcome and how is it measured in the study?

It may be that existing evidence from clinical studies together with already well-established mechanisms is enough to establish the general mechanistic claim. In other cases, the existing evidence fails to establish causality, and it is necessary to identify and evaluate mechanistic studies. To this end, this chapter presents the following five-step strategy for gathering evidence of mechanisms:

1. *Identify*: Identify a number of specific mechanism hypotheses.
2. *Formulate*: For each specific mechanism hypothesis, formulate a number of review questions.
3. *Search*: Use these review questions to search the literature.
4. *Refine*: Identify the evidence most relevant to the mechanism hypothesis by refining the results of this search.
5. *Present*. Present the evidence relevant to the mechanism hypothesis.

This strategy is intended to help overcome some of the practical difficulties with identifying evidence of mechanisms—difficulties which may prevent appraisers from considering all the relevant evidence. Once this evidence of mechanisms has been identified, it can then be evaluated alongside the existing evidence of correlation from clinical studies, as explained in Chaps. 6 and 7.

The overall approach of this chapter is illustrated in Fig. 5.1. The five steps outlined above are explained in detail in the following sections.

## 5.1 Identify Specific Mechanism Hypotheses

**Efficacy.** In order to evaluate the general mechanistic claim that *there is a mechanism that can account for the observed correlation between a putative cause and effect in a study population*, it is useful to identify key features of possible mechanisms

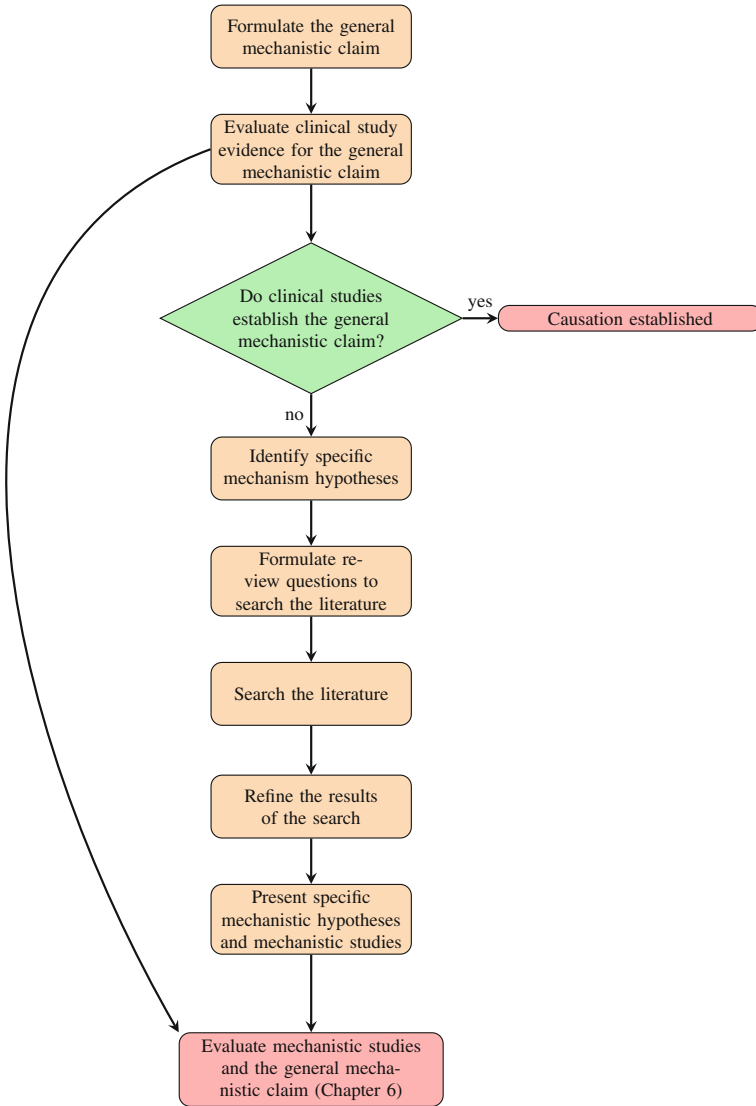


Fig. 5.1 The overall approach to gathering evidence of mechanisms

of action. Each proposed mechanism of action, or partial description of proposed mechanism of action, is a **specific mechanism hypothesis**. But note that a specific mechanism hypothesis need not be a *complete* description of a mechanism.

**Example:** *Specific mechanism hypotheses for determining efficacy.*

Aspirin prevents heart disease via cyclooxygenase (COX) inhibition, and the mechanisms that underlie this prevention are established. However, aspirin also seems to reduce the incidence of some cancers. Here, the mechanisms are much less well understood. As Chan et al. (2011) write: “the mechanism of aspirin’s antineoplastic effect is less clear, with substantial evidence supporting both COX-dependent and COX-independent mechanisms. Moreover, data supporting the importance of COX-dependent mechanisms are not entirely consistent concerning the relative importance of the COX-1 and COX-2 isoforms in carcinogenesis”. In this quotation, the general mechanistic claim is that aspirin exhibits an antineoplastic effect. There are also a couple of more specific mechanism hypotheses, for example, that this antineoplastic effect is mediated by COX-dependent mechanisms. Evidence relating to these more specific mechanism hypotheses provides a way to determine the status of the general mechanistic claim.

**External validity.** In order to evaluate the general mechanistic claim that *there is a mechanism in the target population sufficiently similar to the mechanism responsible for the correlation observed in the study population*, specific mechanism hypotheses need to pertain to the mechanism of action. It is important to consider the possibility that the mechanism in the target population may contain further component mechanisms that counteract the mechanism of action in the study population and affect the extent of the correlation between the putative cause and effect. So one needs to ask, *are there any masking mechanisms in the target population?*

**Example:** *Specific mechanism hypotheses for determining external validity.*

According to NICE guidelines, treatment for hypertension should differ depending on ethnicity (NICE 2011). Although ACE-inhibitors have proved beneficial for hypertension in many study populations, there remains the question of whether they are the optimal treatment in some distinct target population, such as African or Caribbean populations. In this case, it is necessary to determine the status of the following general mechanistic claim: the relevant hypertensive mechanisms in the study populations are sufficiently similar to the mechanisms in African or Caribbean populations. This general mechanistic claim can be evaluated by evaluating a more specific mechanism hypothesis, namely that African and Caribbean populations have a lower renin state. As we shall see in Chap. 6, there is some good mechanistic evidence in favour of this specific mechanism hypothesis, and this undermines the general mechanistic claim. This is why, instead, calcium channel blockers are the recommended antihypertensive treatment in African and Caribbean populations (Clarke et al. 2014).

There are two main ways to identify a specific mechanism hypothesis.

First, a specific mechanism hypothesis may be proposed on the basis of published studies from the clinical study literature. If a clinical study establishes a correlation between a putative cause and effect, and the suggestion is that this correlation is causal, then the authors of such a study usually identify at least one possible mechanism hypothesis of the following form: *It is plausible that mechanism with features F links the putative cause and effect in the study population.* The study may also point out possible masking mechanisms (Illari 2011). Given this, the discussion section of a published paper that reports the results of a clinical study is a good place to look in order to locate a specific mechanism hypothesis.

**Example:** The discussion section of a recent paper on the effect of long-term aspirin use on the risk of cancer says: ‘[O]ur findings suggest that for the gastrointestinal tract, aspirin may influence additional mechanisms critical to early tumorigenesis that may explain the stronger association of aspirin with a lower incidence of gastrointestinal tract cancer. Such mechanisms include modulation of cyclo-oxygenase-2, the principal enzyme that produces proinflammatory prostaglandins, including prostaglandin E2, which increases cellular proliferation, promotes angiogenesis, and increases resistance to apoptosis. Aspirin may also play a role in Wnt signaling, nuclear factor B signaling, polyamine metabolism, and DNA repair’ (Cao et al. 2016). References are given for these specific mechanism hypotheses.

Second, a specific mechanism hypothesis may also be proposed on the basis of existing mechanistic studies or clinical expertise.

**Example:** Large goitres may make it difficult to breathe. It has recently been established that radiotherapy leads to a reduction in the size of large nodular goitres (Nielsen et al. 2006; Bonnema et al. 2007). Will reducing the size of goitres lead to improved respiratory function? Basic clinical experience suggests that there is a mechanism by which a reduction in the size of obstructions in the airway leads to an improvement in respiratory function. This was not established on the basis of clinical studies, but rather on very basic clinical experience. A proponent of this view may propose that this clinical experience supports the existence of a mechanism by which radiotherapy makes a positive difference to respiratory function in patients with large nodular goitres, since large nodular goitres are simply a type of obstruction in the airway that results from an enlargement of the thyroid. However, it may also be proposed that there is a possible masking mechanism. Radiotherapy to the throat might otherwise reduce respiratory function (by, say, causing scarring). A proponent of this view might propose this masking mechanism which may affect the extent of the correlation between radiotherapy and improved respiratory function (Bonnema et al. 2007).

It is important to bear in mind the following practical point. Many policy-makers require an expert evaluation of evidence in their process. For instance, expert evaluations routinely take place at the International Agency for Research on Cancer (IARC), the UK Medicines and Healthcare Products Regulatory Agency (MHRA), the UK National Institute for Health and Care Excellence (NICE), and the EU Committee for Medicinal Products for Human Use (CHMP). In such cases, it may be useful to provide a list of specific mechanism hypotheses to committee members before gathering evidence, in order to give them the opportunity to suggest alterations to the list well in advance of the committee actually meeting (Aronson et al. 2018). Identifying a set of specific mechanism hypotheses at the outset is a good way of proceeding in the face of a large number of mechanistic studies: it makes the process of gathering evidence more manageable by helping to restrict focus to only those published mechanistic studies potentially relevant to the mechanism hypotheses of interest.

## 5.2 Formulate the Review Questions

An effective method for carrying out a review of the literature begins with a well-formulated review question. The suggestion here is to use the specific mechanism hypotheses to help formulate a number of review questions.

Two points are important to keep in mind:

1. Some features of the proposed mechanism may already be established, so it would be unnecessary to look for further evidence in favour of them. Such features should not figure in the more specific review questions. Only the contentious key features of the proposed mechanism should figure in the review question.
2. The review questions may need to be updated in the course of the literature search. In particular, the search may suggest some more specific review questions about the entities, activities, and their organization in the proposed mechanism. Any changes to the review questions should be documented.

**Example:** A number of clinical studies establish that there is a correlation between exposure to benzo[a]pyrene and lung cancer, because exposure to benzo[a]pyrene is correlated with tobacco smoking, which is itself correlated with lung cancer (IARC 2009). But these studies alone were not sufficient to establish causation (IARC 2015). A number of specific mechanism hypotheses might explain the correlation between benzo[a]pyrene and cancer: e.g., (i) The diolepoxide mechanism; (ii) The radical-cation mechanism. These hypotheses lead to the following review questions concerning contentious key features of the respective mechanisms: (i) Do intermediate metabolites of benzo[a]pyrene react with DNA to form DNA adducts associated with tumorigenesis? (ii) Is benzo[a]pyrene oxidized in such a way that leads to free radical formation which may in turn form DNA adducts? These review questions can then be used to search the literature.

The review questions may be formulated according to the PICO framework. PICO stands for Population, Intervention, Comparator, and Outcome (for more information see O'Connor et al. (2011)).

Suppose we are interested in the following research question: *Is there a mechanism in women over fifty linking regularly taking aspirin (rather than not regularly taking aspirin) to developing asthma?* The PICO framework helps in a number of ways to answer this question, by emphasizing what are the most important parts of the research question. Specifically, it picks out the relevant population (women over fifty), the intervention in that population of interest (regularly taking aspirin), and the outcome (developing asthma). It will also identify the comparator (asthma prevalence in members of the same population not regularly taking aspirin). This has the effect of making clear the most important aspects of the intended research objective. In turn, this focuses the search on the most relevant literature, as well as assisting in the presentation of the literature that is obtained by the search.

The PICO framework may be adapted to the research objective at hand. In particular, the PECO framework has been developed for non-interventional studies: Population, Exposure, Comparator, and Outcome (Vandenberg et al. 2016). One can ask, for instance: is there a mechanism in human males (population) linking exposure to high levels of benzo[a]pyrene (exposure) rather than low levels of benzo[a]pyrene (comparator) to scrotal cancer (outcome)?

### 5.3 Search the Literature

A review question can then be used to search the literature for evidence for the contentious key features of a specific mechanism hypothesis. This should take place with the assistance of domain experts.

At this stage, decisions need to be made about which databases and other sources should be searched. These decisions should be documented in order to aid transparency and reproducibility. (See [Appendix A](#) for some examples of databases, Part II for tools to support the process of evidence appraisal, and Sect. 5.6 for a worked example of a literature search.)

One can identify research potentially relevant to the assessment of the specific mechanism hypothesis by looking at the relevant mechanistic study literature:

1. In the first instance, this may be done by following up the references from the discussion section of any clinical study report which proposes a mechanism as the best explanation of an observed correlation. Any other publicly available reports may be useful here also, e.g., government agency reports, doctoral theses, etc.
2. More systematically, a preferred method for searching the literature may be used, e.g., a [PubMed](#) search using appropriate Medical Subject Heading (MeSH) terms, including key terms from the hypothesized mechanisms.

Efforts to standardise terminology and indexing practices for publications reporting mechanistic studies are welcome, especially in order to facilitate text mining techniques, which are becoming increasingly widespread. It is also important that even the negative findings of mechanistic studies are published, to reduce publication bias.

## 5.4 Refine Results of the Search

Identifying evidence from the literature requires expert judgement, which is susceptible to bias. In order to guard against the effects of such biases, the details of the search procedure should be clearly presented (O'Connor et al. 2011). This protects against the effects of bias by providing a transparent and reproducible literature search strategy (Vandenberg et al. 2016).

A study flow diagram can be used to present the process of selecting studies for inclusion in the review (O'Connor et al. 2011). This can be made with reference to the guidance in the PRISMA framework (Moher et al. 2009). According to this guidance, a study flow diagram consists of four phases: Identification, Screening, Eligibility, and Inclusion. After identifying studies by searching databases with a review question, the studies are then screened for duplicates, and excluded studies are recorded. The eligibility of the studies is then determined, and any ineligible studies are recorded as excluded along with the reasons for their exclusion. This leaves the included studies.

A key question here is: *Is any of this evidence not relevant?*

1. Use preferred inclusion and exclusion criteria and expert knowledge to rule out irrelevant mechanistic studies (Kushman et al. 2013).
  - *Does the publication include original data?* A good rule of thumb: if it does not include original data, then exclude the publication.
2. It may be possible to exclude some studies by a review of the title and abstract. A full-text review may be necessary to exclude other studies.
  - All excluded studies should be documented, along with the reasons for exclusion.
3. There are content management tools available to help in identifying, screening, organizing, and summarizing the evidence.
  - For example: Health Assessment Workspace Collaborative (HAWC). See: <https://hawcproject.org/>.

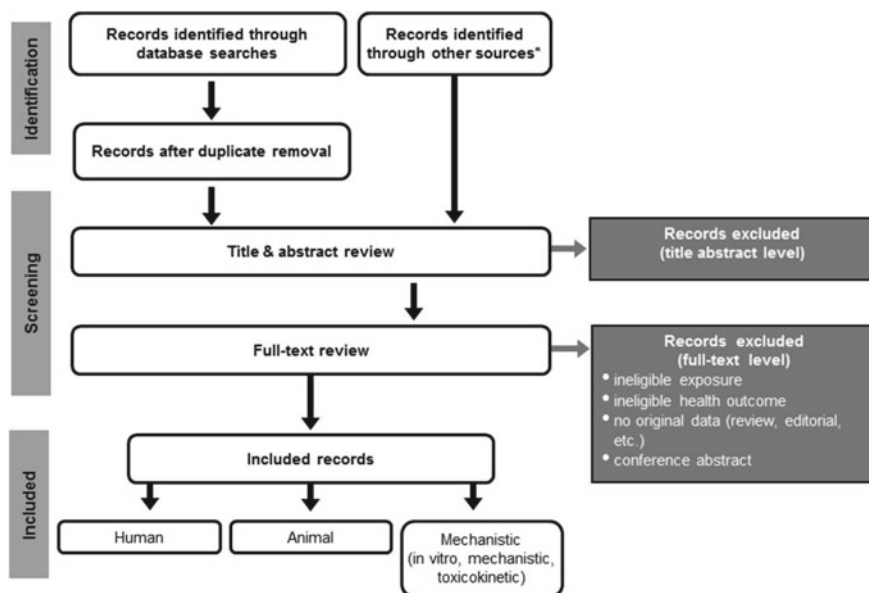


Fig. 5.2 An example study flow diagram reproduced from Vandenberg et al. (2016)

An example study flow diagram for evidence of mechanisms is presented in Fig. 5.2 (Vandenberg et al. 2016).

## 5.5 Presenting the Evidence of Mechanisms

A clear summary of the identified evidence of mechanisms is an important precursor to evaluating that evidence. (Presenting the quality of evidence of mechanisms is a separate issue, for which guidance is provided in Sect. 6.4.) A summary of evidence of mechanisms should clearly state the general mechanistic claim that the mechanism in question is proposed to account for, that is, whether it is presented as evidence of the existence of a mechanism of action for efficacy, or as evidence of similarity of mechanisms between populations to account for external validity. This includes a clear statement of the cause *A* under investigation as well as the particular outcome *B* of interest. The presentation of evidence should also make clear the specific mechanism hypotheses under consideration, and present the evidence in favour of the contentious key features of the specific mechanism hypotheses.

**Example:** *IARC's overall process of gathering and presenting evidence of mechanisms.*

In order to help identify and organise further evidence of mechanisms in the literature, the International Agency for Research on Cancer makes use of existing evidence of mechanisms in the form of ten key characteristics, one or more of which are frequently exhibited by known carcinogens (Smith et al. 2016). In our terminology, the ten key characteristics are key features of specific mechanism hypotheses, which are possible instantiations of the general mechanistic claim that there is a mechanism linking the considered exposure to cancer in the relevant sites in humans. The ten key characteristics are the ability of the putative carcinogen to:

1. Act as an electrophile either directly or after metabolic activation;
2. Be genotoxic;
3. Alter DNA repair or cause genomic instability;
4. Induce epigenetic alterations;
5. Induce oxidative stress;
6. Induce chronic inflammation;
7. Be immunosuppressive;
8. Modulate receptor-mediated effects;
9. Cause immortalization;
10. Alter cell proliferation, cell death, or nutrient supply.

For instance, a correlation between benzene and cancer in humans has been observed in many studies. In order to determine whether this correlation is causal, it is necessary to determine the status of the relevant general mechanistic claim, namely, that there exists a mechanism linking exposure to benzene to cancer in humans that can account for the extent of the observed correlation (IARC 2015). A first step is to propose specific mechanism hypotheses, with the help of the ten key characteristics. For example, the specific mechanism hypothesis might be that benzene induces certain chromosomal aberrations that are characteristic of carcinogens. This leads to review questions that help to identify evidence relevant to this specific mechanism hypothesis. In this case, there is mechanistic evidence that exposure to benzene causes chromosomal aberrations in vivo in bone marrow cells of mice and rats. There is also mechanistic evidence that benzene exposure also causes chromosomal aberrations and mutation in human cells in vitro. This mechanistic evidence should be listed alongside the specific mechanism hypothesis and will adjudicate on the contentious features of the proposed mechanism. The identified evidence may be sufficient to determine the status of the general mechanistic claim, but this would involve first evaluating the evidence of mechanisms, which is the topic of Chap. 6.

## 5.6 Worked Example on Probiotics and Dental Caries

This worked example shows how our general method for gathering evidence of mechanisms can be applied to a specific case dealing with the effectiveness of probiotics for dental caries.

**Identify specific mechanism hypotheses for probiotics in preventing dental caries.** Cagetti et al. (2013) conducted a review of the caries-prevention effect of probiotics in human. Three studies were found assessing caries lesion development as outcome, with a further 20 studies reporting only caries risk factors as interim outcomes. The authors concluded "...[t]he effect of probiotics on the development of caries lesion seems encouraging, but to date, RCTs on this topic are insufficient to provide scientific clinical evidence."

More recently, a systematic review on probiotics and oral health (Seminario-Amez et al. 2017) reached similar conclusions on the effectiveness for the prevention of dental caries; laboratory data and the effect on interim outcomes is promising, but long-term clinical trials are needed.

In the review by Cagetti et al. (2013), the mechanisms of action of probiotics were described. These were:

- adhesion
- co-aggregation
- competitive inhibition
- production of organic acids
- bacteriocin-like compounds
- immune-modulation (Teughels et al. 2008)

Cagetti et al. (2013) did note that not all of these mechanisms were fully understood. Seminario-Amez et al. (2017) also noted that the mechanism of probiotics in the oral cavity is not clearly established. However, studies are cited to support the role of probiotics in reducing counts of cariogenic pathogens, inhibiting periodontal pathogens, modulating the inflammatory response and producing beneficial substances. The ability of probiotics to compete with pathogens for adhesion surfaces and nutrients, causing displacement of the latter ones, was also confirmed in laboratory studies.

**Formulate the review questions and search the literature.** In order to further explore how probiotics might work for the prevention of dental caries, we searched for review articles describing the mechanism of action. Four relevant articles were found (Bonifait et al. 2009; Caglar et al. 2005; Saha et al. 2012; Singh et al. 2013).

**Refine results of the search.** Bonifait et al. (2009) postulated that "[t]o have a beneficial effect in limiting or preventing dental caries, a probiotic must be able to adhere to dental surfaces and integrate into the bacterial communities making up the dental biofilm. It must also compete with and antagonize the cariogenic bacteria and thus prevent their proliferation. Finally, metabolism of food-grade sugars by the probiotic should result in low acid production." Bonifait et al. (2009) cite a number of

studies showing the different abilities of the probiotics, such as the ability to integrate with the biofilm, and conclude that probiotics can neutralize acidic conditions in the mouth and interfere with cariogenic bacteria. The same evidence is cited in Singh et al. (2013).

**Present the evidence of mechanisms.** The number of studies investigating the effectiveness of probiotics for the prevention of dental caries is limited. There is a body of evidence from laboratory studies and clinical trials that interim outcomes linked with reduced dental caries can be improved through the use of probiotics. Several specific mechanism hypotheses were found in this research, mainly dealing with local (rather than systemic) effects of probiotics. However, not all mechanisms are yet fully understood.

In this example, understanding how probiotics might work through the various mechanisms of action helps to interpret the limited evidence of effectiveness. Probiotics are likely to have a preventive effect on dental caries, effected through a range of known mechanisms. Probiotics are also very unlikely to have significant adverse effects (Borriello et al. 2003).

We did not undertake a systematic review of the evidence on how probiotics might work. However, there appears to be a consistent view of the underlying mechanisms between the publications reviewed here. In this case, where unintended consequences are likely to be minimal due to the already wide and safe use of probiotics, a systematic review may not be needed to generate evidence of mechanisms.

## References

- Aronson, J. K., La Caze, A., Kelly, M. P., Parkkinen, V.-P., & Williamson, J. (2018). The use of mechanistic evidence in drug approval. *Journal of Evaluation in Clinical Practice*. <https://doi.org/10.1111/jep.12960>
- Bonifait, L., Chandad, F., & Grenier, D. (2009). Probiotics for oral health: Myth or reality? *Journal of the Canadian Dental Association*, 75(8), 585–590.
- Bonnema, S. J., Nielsen, V. E., Boel-Jorgensen, H., Grupe, P., Andersen, P. B., Bastholt, L., et al. (2007). Improvement of goiter volume reduction after 0.3 mg recombinant human thyrotropin-stimulated radioiodine therapy in patients with a very large goiter: a double-blinded, randomized trial. *The Journal of Clinical Endocrinology and Metabolism*, 92(9), 3424–3428.
- Borriello, S., Hammes, W., Holzapfel, W., Marteau, P., Schrezenmeir, J., Vaara, M., et al. (2003). Safety of probiotics that contain lactobacilli or bifidobacteria. *Clinical Infectious Diseases*, 36(6), 775–780.
- Cagetti, M. G., Mastroberardino, S., Milia, E., Cocco, F., Lingström, P., & Campus, G. (2013). The use of probiotic strains in caries prevention: A systematic review. *Nutrients*, 5(7), 2530–2550.
- Caglar, E., Kargul, B., & Tanboga, I. (2005). Bacteriotherapy and probiotics' role on oral health. *Oral diseases*, 11(3), 131–137.
- Cao, Y., Nishihara, R., Wu, K., Wang, M., Ogino, S., Willett, W., et al. (2016). Population-wide impact of long-term use of aspirin and the risk for cancer. *JAMA Oncology*, 2(6), 762–769.
- Chan, A., Arber, N., Burn, J., Chia, W., Elwood, P., Hull, M., et al. (2011). Aspirin in the chemoprevention of colorectal neoplasia. *Cancer Prevention Research*, 5(2), 164–78.
- Clarke, B., Gillies, D., Illari, P., Russo, F., & Williamson, J. (2014). Mechanisms and the evidence hierarchy. *Topoi*, 33(2), 339–360.

- IARC (2009). *IARC Monographs on the evaluation of carcinogenic risks to humans, vol. 100F: A review of human carcinogens. Chemical agents and related occupations*. Lyon: International Agency for Research on Cancer.
- IARC (2015). *IARC Monographs on the evaluation of carcinogenic risks to humans: Preamble*. Lyon: International Agency for Research on Cancer.
- Illari, P. M. (2011). Mechanistic evidence: Disambiguating the Russo-Williamson thesis. *International Studies in the Philosophy of Science*, 25(2), 139–157.
- Kushman, M., Kraft, A., Guyton, K., Chiu, W., Makris, S., & Rusyn, I. (2013). A systematic approach for identifying and presenting mechanistic evidence in human health assessments. *Regulatory Toxicology and Pharmacology*, 67(2), 266–277.
- Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Med*, 6(7), e1000097.
- NICE. (2011). *CG127: Hypertension: Full guideline*. London: National Institute for Health and Clinical Excellence.
- Nielsen, V. E., Bonnema, S. J., Boel-Jorgensen, H., Grupe, P., & Hegedüs, L. (2006). Stimulation with 0.3-mg recombinant human thyrotropin prior to iodine 131 therapy to improve the size reduction of benign nontoxic nodular goiter: a prospective randomized double-blind trial. *Archives of Internal Medicine*, 166(14), 1476–1482.
- O'Connor, D., Green, S., & Higgins, J. (2011). chapter 5: Defining the review question and developing criteria for including studies. In *Cochrane handbook of systematic reviews of intervention*. The Cochrane Collaboration. Version 5.1.0 (updated March 2011).
- Saha, S., Tomaro-Duchesneau, C., Tabrizian, M., & Prakash, S. (2012). Probiotics as oral health biotherapeutics. *Expert opinion on biological therapy*, 12(9), 1207–1220.
- Seminario-Amez, M., López-López, J., Estrugo-Devesa, A., Ayuso-Montero, R., & Jané-Salas, E. (2017). Probiotics and oral health: A systematic review. *Medicina oral, patología oral y cirugía bucal*, 22(3), e282–e288.
- Singh, V. P., Sharma, J., Babu, S., Singla, A., et al. (2013). Role of probiotics in health and disease: A review. *JPMA. The Journal of the Pakistan Medical Association*, 63(2), 253–257.
- Smith, M. T., Guyton, K. Z., Gibbons, C. F., Fritz, J. M., Portier, C. J., Rusyn, I., et al. (2016). Key characteristics of carcinogens as a basis for organizing data on mechanisms of carcinogenesis. *Environmental Health Perspectives*, 124, 713–21.
- Teughels, W., Van Essche, M., Sliepen, I., & Quirynen, M. (2008). Probiotics and oral healthcare. *Periodontology 2000*, 48(1), 111–147.
- Vandenberg, L., Agerstrand, M., Beronius, A., Beausoleil, C., Bergman, A., Bero, L., et al. (2016). A proposed framework for the systematic review and integrated assessment (SYRINA) of endocrine disrupting chemicals. *Environmental Health*, 15(74), 1–19.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



# Chapter 6

## Evaluating Evidence of Mechanisms



**Abstract** In this chapter, we discuss how to evaluate evidence of mechanisms. This begins with an account of how a mechanistic study provides evidence for features of specific mechanism hypotheses, laying out a three step procedure of evaluating: (1) the methods used, (2) the implementation of the methods, and (3), the stability of the results. The next step is to combine those evaluations to present the quality of evidence of the general mechanistic claim.

Having explained how evidence of mechanisms can be obtained, the next step is to evaluate that evidence, which is the topic of this chapter. In the following chapter will explain how this evaluation can be integrated with an evaluation of evidence for a correlation in order to determine an overall evaluation of the causal claim of interest.

### 6.1 Overview

Evaluating evidence of mechanisms should start with clear formulations of the [general mechanistic claim](#) and each [specific mechanism hypothesis](#), for which evidence is gathered via the procedure described in Chap. 5. The general mechanistic claim concerns either the existence of a mechanism (to account for efficacy) or the similarity of mechanisms between populations (to account for external validity). The specific mechanism hypotheses posit key features of potential mechanisms of action; corroborating evidence for the specific mechanism hypotheses thus supports the general mechanism claim.

Evaluating evidence of mechanisms requires assessing the reliability of the methods and techniques by which the evidence was produced. For a general mechanistic claim about the existence of a mechanism, this evidence may come from clinical studies that report a strong correlation between variables. Clinical study evidence should be evaluated according to normal criteria of good experimental design and analysis—see, e.g., Chow and Liu (2004). However, a mere correlation, even a strong one, may result from unmeasured confounding factors. Thus, only when clinical

study evidence is high quality can it significantly support a claim about the existence of a mechanism. Similarly, observing a clear dose-response relationship between variables can lend credibility to a causal interpretation (Hill 1965), and thus to the existence of a linking mechanism. Note, however, that biological mechanisms often exhibit feedback regulation and other complex behaviours that do not give rise to clear dose-response relationships. The lack of a dose-response relationship is thus not strong evidence against the existence of a mechanism. For establishing similarity of mechanisms, one normally needs some evidence of the details of the specific features of the relevant mechanisms.

A [mechanistic study](#) provides evidence for features of specific mechanism hypotheses. Mechanistic studies are conducted by one or more of the following three means:

1. **Experimental manipulation:** by finding a suitable experimental system in which the mechanism or parts of it are present, making predictions about the mechanism's behaviour under interventions on some of its parts, and comparing the predictions to the outcomes of experiments where those parts are actually manipulated. Standard tools for evaluating the quality of experimental design, data analysis, randomisation procedure (when applicable) and statistical inference can thus be applied to evaluate the possibility of experimental error (Montgomery 2009). Simulation experiments can also be used, especially to investigate whether the hypothesised organisation of a mechanism is in fact sufficient for producing the phenomenon of interest. However, the modelling assumptions on which a simulation is based should be corroborated by empirical evidence before the results of a simulation can be considered as evidence for causal claims.
2. **Observation:** entities, activities and organisation of a mechanism can be found by observation techniques such as imaging technologies, autopsy, (molecular) epidemiological studies, and social surveys (for mechanisms that include parts of the social environment as components, or which are sensitive to sociological variables like socioeconomic status, parental or neighbourhood effects).
3. **Analogy:** Sometimes a mechanism can be hypothesised, and, to a low degree, even confirmed, by analogy to an established mechanism linking a closely similar intervention/exposure to a similar outcome.

The particular challenges for evaluating evidence for features of mechanisms stem from the fact that the evidence is often produced in systems in which most of the natural context of the mechanism is absent (e.g., in vitro studies), or in which the context and possibly the mechanism itself is different from humans (e.g., model organism studies). Model organism studies are susceptible to bias in the same way as human trials. Standard ways of evaluating statistical errors or bias due to trial design may be used to assess the quality of trials conducted on experimental animals (Chow and Liu 2004). In the case of in vitro studies that require extensive preparation of samples and employ complicated and indirect detection methods, there is always the risk that an experimental result is an artefact produced by the instruments or

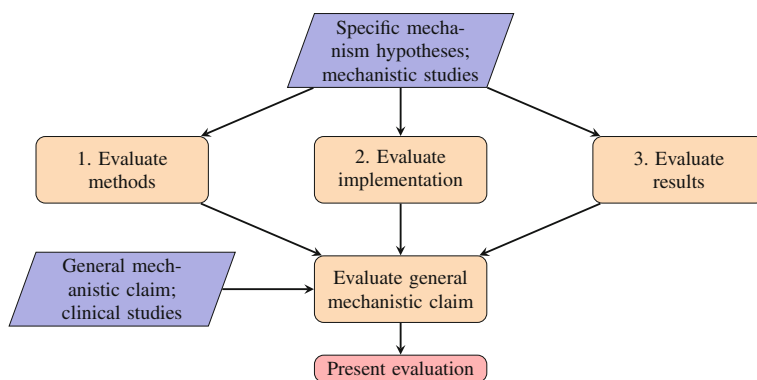
preparation methods, rather than a feature belonging to the actual mechanism. In addition to evaluating the possibility of mere experimental error and bias, weighing evidence of mechanisms requires evaluating how well these problems have been mitigated in the process of creating the evidence.

Below we describe a procedure for evaluating evidence from mechanistic studies, broken down to three steps:

1. Evaluating the methods used,
2. Evaluating the implementation of the methods, and
3. Evaluating the stability of the results.

Each step involves evaluating the mechanistic studies by means of particular quality indicators. Evidence that ranks well (respectively, badly) in the light of several indicators ought to be taken as higher (respectively, lower) quality than evidence that ranks well (respectively, badly) with respect to fewer considerations. Note that this is not a rigidly algorithmic approach. Instead, domain-specific expertise should be employed in interpreting results and must be allowed to adjust the overall quality ranking. There are also trade-offs between the quality indicators; these are pointed out below. Finally, in cases where one has evidence that supports the general mechanistic claim directly, e.g. a high quality clinical trial, as well as evidence in support of some specific mechanism hypotheses (see Fig. 3.1), one needs to combine these to come up with a final quality status for the general mechanistic claim.

The procedure of this section is summarised in Fig. 6.1. The three-step method for evaluating mechanistic studies is presented in the next section, Sect. 6.2. These steps contribute to the evaluation of the general mechanistic claim as described in Sect. 6.3. Finally, Sect. 6.4 describes how the evaluation of evidence of mechanisms can be presented.



**Fig. 6.1** A procedure for evaluating evidence of mechanisms

## 6.2 Evaluating Mechanistic Studies

This section further develops the three-step procedure outlined above.

**Step 1. Evaluate methods.** The first step is to evaluate the methods employed by the studies under review. Methods should be evaluated with respect to their typical error characteristics. This requires an amount of domain specific expert knowledge, but typically there are some paradigmatic examples of well conducted studies and reliable methods that can serve as a benchmark for evaluating the reliability of methods. A precondition for evaluating methods is that the methods themselves and their error characteristics are understood. This gives us three general quality indicators, described below.

1. *Well understood methods and model systems.* In order to evaluate mechanistic studies as high quality, it is normally essential to establish that the methods by which the evidence was produced are reliable. The better one understands how a method works, the easier it is to evaluate its reliability. Understanding how a method works is thus normally a precondition for attributing high quality to an item of evidence produced by that method. This applies to experimental model systems as well. Evidence produced in well understood model systems, in which the mechanisms responsible for the experimental result can be directly compared to relevant mechanisms in humans, should be given higher credence than evidence produced in model systems whose functioning is poorly understood. This indicator trades off against indicator (2) below: well characterised and understood experimental systems are typically simple, and thus often fail to faithfully reflect the whole-organism level physiology of humans.
2. *The degree to which experimental systems replicate human features of interest, and the quality of experimental animal trials.* Model systems that faithfully replicate human features of interest have greater external validity than ones that are very dissimilar to humans. The greater the similarity between an experimental model system and humans, the higher the quality of the evidence gleaned from the model. Notice a trade off between the choice of a model by its similarity to humans, and the tractability of the model itself. The most well understood experimental models are typically highly dissimilar to humans, whereas models that faithfully replicate many features of humans are considerably less well understood on the whole. Models that are very well characterised, but highly dissimilar to humans, are often used in basic science research that aims to discover highly general mechanisms potentially shared across many species, and such models are indispensable for this purpose. However, when the main focus of research is on justifying claims about causality in humans, the similarity of model systems to humans is an important consideration to keep in mind in evaluating evidence obtained in diverse experimental systems. This indicator trades off against indicator (1), as explained above. Studies performed on experimental animals may offer more conclusive evidence of the operation of an underlying mechanism, as more invasive intervention and measurement methods may be

used in experimental animals than in humans. Animal trials are susceptible to bias in the same way as human studies, and should be evaluated similarly.

3. *The appropriateness of surrogate endpoints.* In some cases, it is not straightforward to directly measure an outcome of interest. However, it may be possible to measure some distinct endpoint as a way of indirectly measuring the endpoint of interest. Such a distinct endpoint is sometimes called a surrogate endpoint. For example, blood pressure may be used as a surrogate endpoint for left ventricular function, since it is more straightforward to directly measure blood pressure than left ventricular function, say, by echocardiography (Aronson 2005). Crucially, an endpoint is more likely to be an informative surrogate for the endpoint of interest if it features in the mechanism productive of that endpoint of interest. For example, there is a mechanism linking elevated cholesterol to an increase in the risk of heart disease, and so cholesterol levels are often used as a surrogate endpoint for risk of heart disease. As a result, evaluating evidence of mechanisms is important for the validation of surrogate endpoints (AHRQ 2013). Indeed, in some cases overlooking mechanistic evidence has led to an inappropriate choice of surrogate endpoints and harmful consequences, for example, the recommendation of anti-arrhythmic drugs on the basis of employing ventricular ectopic beat as a surrogate endpoint for cardiac mortality (Holman 2017).

**Step 2. Evaluate implementation.** The second step is to evaluate how well the individual studies have implemented the methods used. Different methods have their typical error characteristics. For instance, trials may produce biased results if randomisation is not implemented appropriately, or imaging technologies may produce artefacts. Assessing the implementation of methods consists in evaluating what means have been taken to control for the characteristic errors of the study methods. Doing this requires some knowledge of the typical error characteristics of different methods. One should thus consider the quality indicator (1) first: if the principles of operation of a particular method are poorly understood, it is more likely that one fails to distinguish and control for experimental artefacts and biased results. After that, one should assess whether the methods were implemented with appropriate precautions to control for known error types. It is typically impossible to ensure that all possible sources of error have been controlled for in implementing a particular method.

**Step 3. Evaluate results.** The third step is to evaluate the stability of the results. High credence in the validity of a result can be conferred by finding that several independent methods provide similar results. This is an important indicator of the reliability of a result:

4. *Independent detectability.* The greater the number of independent methods that are able to confirm features of a mechanism, the more confident one can be that the observations are real and not artefacts.

However, one should also assess whether results are consistent across studies conducted in similar settings using similar methods. This gives us a further quality indicator:

5. *Consistency*. Inconsistencies that cannot be explained as resulting from differences in methods or relevant contextual factors, or as resulting from poor implementation of methods in some of the studies, should result in lowering the quality status of the evidence.

Finally, one should assess how tolerant the confirmed mechanisms are to variation in background conditions or properties of the parts of the mechanism itself. Mechanisms that are highly robust in the sense that their operation is not disturbed by such variation are more likely to be extrapolatable between heterogeneous contexts than mechanisms that are sensitive to such variation.

6. *Robustness of features across varying contexts*. The greater the variability of contexts or model systems in which some or all features of a mechanism are found, the more plausible it is that the results are extrapolatable. This may be understood as application of Hill's consistency indicator to evidence of mechanisms (Hill 1965).

### 6.3 Determining the Status of the General Mechanistic Claim

This section describes how the [status](#) of the general mechanistic claim can be assessed, based on the evaluation of the mechanistic study evidence for the specific mechanism hypotheses and the evaluation of the clinical study evidence for the general mechanistic claim.

Recall that different types of general mechanistic claim need to be considered for the purpose of evaluating efficacy and for the purpose of evaluating external validity. In the former case, one considers the question of whether there is a mechanism capable of accounting for the observed correlation. In the latter case, one considers the similarity of mechanisms between the study and the target populations. The two boxes below describe typical conditions in which one would attribute a high (or low) status to either type of general mechanistic claim. As evidence of mechanisms can be highly heterogeneous, these conditions should not be thought of as exhaustive, nor as giving a mechanical procedure for attributing status. Instead, they are to be thought of as heuristics that need to be considered in the light of relevant domain-specific expertise, to arrive at a decision about the status of the general mechanistic claim (see also the tools in Chap. 4).

**Checklist of questions to consider in evaluating a general mechanistic claim for efficacy**

*Does the evidence warrant conferring a higher status to a mechanistic existence claim?* Consider the following questions about the evidence; can one or more be answered in the affirmative?

1. Has a correlation of the same size been established in many studies under slightly varying circumstances (robust detectability)? If yes, is it likely that the population of interest falls within the range of circumstances which have been tested?
2. Is the observed correlation so large that it is very unlikely to be explained by bias or confounding, leaving the existence of a mediating mechanism as the most plausible explanation?
3. Is the mechanism known in some detail? Can it account for the correlation and its size? Are most of the crucial features of the mechanism known and understood? Does the mechanism support novel predictions?
4. Is it plausible that the behaviour of the mechanism crucially depends on just some components or organisational features? If so, are such critical features well established according to the considerations described above? This can provide sufficient grounds for assigning the mechanistic claim a higher status than it would otherwise have. Example: consider a biochemical pathway with a single rate-limiting step. In such a case, establishing the rate-limiting step is usually more important for understanding the behaviour of the whole mechanism than establishing the rate of the reactions downstream from that step.

*Does the evidence warrant conferring a lower status to a mechanistic existence claim?* Consider the following questions about the evidence; can one or more be answered in the affirmative?

1. Is a counteracting mechanism likely? If so, could the correlation the mechanism is posited to explain be spurious? (If the existence of a mechanism is inferred from clinical studies, discovering that the observed correlation might be spurious counts as evidence against existence of the purported underlying mechanism as well.) If the evidence does not suggest that the correlation is spurious, this does not mean that one should revise the conclusion about the existence of a mechanism. Rather, evidence of masking suggests that the (masked) mechanism will not reliably support efficacious interventions unless the masking mechanisms can be controlled for.
2. Does the mechanism exhibit such complexity that its overall behaviour is very unpredictable?
3. Is the hypothesised mechanism inferred from evidence of an analogous mechanism or mechanisms in some other domain?

**Checklist of questions to consider in evaluating a general mechanistic claim for external validity**

*Does the evidence warrant conferring a higher status to a mechanistic similarity claim?* Consider the following questions about the evidence; can one or more be answered in the affirmative?

1. Has a correlation of the same size been established in several studies under slightly varying circumstances (robust detectability), and in several populations that are related to the target population (e.g., phylogenetically, geographically), in such a way that these correlations cannot be explained by bias or confounding, and one must posit a similar mechanism operating in all the populations to explain the observed correlations?
2. Is the mechanism known in some detail both in the study population and the target population, and found to be similar in both, and such that it can account for the observed correlation? This can be established by applying the considerations described above.
3. When the behaviour of the whole mechanism crucially depends on some component(s) or an organisational feature, are the critical features of the mechanism similar in the study and the target populations? If so, this can provide sufficient grounds for assigning the mechanistic claim a higher status than it would otherwise have.

*Does the evidence warrant conferring a lower status to a mechanistic similarity claim?* Consider the following questions about the evidence; can one or more be answered in the affirmative?

1. Is a counteracting mechanism in the target population likely? Does this suggest that the correlation that the mechanism is posited to explain is spurious? If not, this does not mean that one should revise the conclusion about the existence of a mechanism. Rather, evidence of masking suggests that the (masked) mechanism will not reliably support efficacious interventions unless the masking mechanisms can be controlled for.
2. Is there dissimilarity between the mechanisms in the study and the target populations?
3. Does the mechanism proposed to support external validity exhibit such complexity that its overall behaviour is unpredictable?
4. Are the hypothesised mechanisms inferred from evidence of an analogous mechanism or mechanisms in some other domain?

Mechanistic evidence for efficacy or external validity should be evaluated considering the correlational evidence that it is invoked to explain. There may be cases in which one has good evidence of mechanisms from analytical studies—e.g., from bench research on experimental systems—that could be invoked to explain a particular correlation, but the correlation in question is not itself well established. This

**Table 6.1** Determining the status of the general mechanistic claim (GMC) on the basis of evidence from mechanistic studies and from clinical studies

		Status of the GMC on the basis of mechanistic studies						
		Established	Provisionally established	Arguable	Speculative	Arguably false	Provisionally ruled out	Ruled out
Status of GMC on basis of clinical studies	Established	Established	Established	Established	Established	Established	Provisionally established	Speculative
	Provisionally established	Established	Established	Provisionally established	Provisionally established	Arguable	Speculative	Arguably false
	Arguable	Established	Provisionally established	Provisionally established	Arguable	Speculative	Arguably false	Ruled out
	Speculative	Established	Provisionally established	Arguable	Speculative	Arguably false	Provisionally ruled out	Ruled out
	Arguably false	Established	Arguable	Speculative	Arguably false	Provisionally ruled out	Provisionally ruled out	Ruled out
	Provisionally ruled out	Arguable	Speculative	Arguably false	Provisionally ruled out	Provisionally ruled out	Ruled out	Ruled out
	Ruled out	Speculative	Arguably false	Ruled out	Ruled out	Ruled out	Ruled out	Ruled out
	Ruled out	Speculative	Arguably false	Ruled out	Ruled out	Ruled out	Ruled out	Ruled out

suggests that there could be hitherto unidentified masking mechanisms that interfere with the operation of the mechanism of interest, or that the mechanism might exhibit stochastic behaviour that does not manifest as an easily detectable correlation. Such considerations should be taken into account in assessing the status of a general mechanistic claim. In evaluating a general mechanistic claim, evidence arising from clinical studies and evidence arising from mechanistic studies have mutually supporting roles.

Table 6.1 determines the status of the general mechanistic claim given the status of the general mechanistic claim based on only clinical studies and its status based on only mechanistic studies. This highlights the mutually supporting roles of mechanistic studies and clinical studies. Note, finally, that determining the status of the general mechanistic claim by combining evidence from clinical and mechanistic studies should not be confused with the task of determining the status of the causal claim on the basis of the status of the general mechanistic claim and the status of the correlational claim—a point which is discussed further at the end of Sect. 7.1 when we develop the analogy of reinforced concrete.

### 6.4 Presenting the Quality of Evidence of Mechanisms

Preparing and presenting summaries of the quality of mechanistic evidence in a standardised manner can be challenging, as evidence of mechanisms comes from highly heterogeneous sources and may involve a mixture of quantitative and qualitative rela-

tionships. Some general guidance can nonetheless be given. The following questions need to be addressed when presenting the status of the general mechanistic claim.

**Presenting the status of the general mechanistic claim for efficacy.** The following questions should be addressed:

1. What is the intervention or exposure level?
2. What is the outcome and how is it measured?
3. What is the status of the general mechanistic claim? Questions to be considered here are, for instance (see Sects. 6.2 and 6.3): Does the clinical study evidence make the general mechanistic claim plausible? What are the specific mechanism hypotheses? Are there any serious gaps in the evidence for these claims? Are there any serious inconsistencies in the evidence for these claims? Is there any serious indirectness (see Sect. 4.6)? Is counteracting plausible?

**Presenting the status of the general mechanistic claim for external validity.** The following questions should be addressed:

1. What is the target population?
2. What is the study population?
3. What is the intervention or exposure level in the target?
4. What is the outcome and how is it measured in the target?
5. What is the intervention or exposure level in the study?
6. What is the outcome and how is it measured in the study?
7. What is the status of the general mechanistic claim concerning similarity? Questions to be considered here are, for instance (see Sects. 6.2 and 6.3): What is the hypothesised mechanism in the study population? Are there any serious gaps in the evidence? Are there any serious inconsistencies in the evidence? Is there any serious indirectness? Is counteracting plausible? Is there any phylogenetic evidence? Is the evidence robust?

When presenting the status of a specific mechanism hypothesis, the quality of the overall evidence of a mechanism should be presented in such a way that it also outlines the quality of the evidence for each of the individual component features of

the mechanism, evaluated by employing the considerations for evaluating evidence described in Sect. 6.2. For example, suppose that a drug is hypothesised to work by binding to a particular receptor on a particular type of cell. The quality of the evidence for this interaction within the overall mechanism should be evaluated by assessing the studies providing evidence for the structure of both the drug and the receptor type, as well as any direct evidence estimating the binding affinity of the drug to its intended target. The greater the number of independent studies, employing well-established experimental methods that are able to confirm the hypothesised interaction, the higher the quality of evidence for this particular feature of the hypothesised mechanism. Conversely, if the evidence for particular features of a mechanism is inconsistent, or gleaned from few studies known to be susceptible to bias, the quality of evidence for those features of the mechanism should be considered low.

To indicate the status of particular features of the mechanism, and the general mechanism claim, one can use the following symbols:

Status	Symbol
Established	*
Provisionally established	++
Arguable	+
Speculative	?
Arguably false	-
Provisionally ruled out	--
Ruled out	#

A brief verbal explanation can be included, e.g. ++; inconsistencies. These symbols can be added to a diagram of a specific mechanism hypothesis, in order to represent the status of key features of the mechanism.

For a critical appraisal tool for mechanistic evidence which summarises key aspects of the evidence gathering process described in Chap. 5, and the evaluation process outlined in this section, see Sect. 4.5.

This system of evaluating and summarizing evidence is not meant as a replacement for other well established evidence assessment frameworks such as GRADE. Rather, the considerations outlined here can often be integrated to existing approaches. For an example of how some of these considerations may be incorporated into the popular GRADE system by a simple amendment of the GRADE evidence profile tables, see Sect. 4.6. Our other tools in Chap. 4 also demonstrate how the evaluation of evidence of mechanisms can be integrated into existing evidence appraisal practices.

**Example:** *ACE inhibitors.*

ACE inhibitors work by modulating the functioning of the renin-angiotensin system (RAS), which is involved in regulation of the sodium concentration of blood, and arterial blood pressure. The basic architecture of RAS regarding blood pressure regulation has been corroborated by numerous studies employing varying methods—see, e.g., Fyhrquist and Saijonmaa (2008) for a review. Thus, there are no particularly contentious parts that would necessitate an in-depth evaluation of the evidence, earning the specific mechanism hypothesis a status of established (indicated by \*). This suffices to establish the general mechanistic claim in support of efficacy in those populations in which trial evidence shows a correlation between ACE inhibitor treatment and blood pressure lowering. To establish the external validity of the blood pressure lowering effect of ACE inhibitors, one needs to establish the general mechanistic claim stating that the RAS mechanisms in the study and the target populations are similar enough.

However, evidence from two subgroup analyses of the ALLHAT (Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial) trial suggested that there were difficulties in establishing efficacy for ACE inhibitors in African Caribbean populations. Piller et al. (2006) showed much higher rates of angioedema (an important and serious side-effect of ACE inhibitor treatment) in African Caribbean individuals, while Leenen et al. (2006) showed that calcium channel blockers (CCB) showed better efficacy than ACEi in that population. The key component of the mechanism regarding the efficacy of ACE inhibitors in African Caribbean populations is renin—an enzyme involved in the production of angiotensinogen, which is further converted by ACE into angiotensin I, and angiotensin II, a highly potent vasoconstrictor. Inhibiting ACE leads to downregulation of angiotensin II, thus inhibiting the RAS mechanism from increasing blood pressure. Low level of renin activity makes the ACE inhibitors much less effective as means to control RAS functioning. There is high quality mechanistic evidence that the African Caribbean population is characterised by low renin profile (Khan and Beevers 2005). There is thus high quality evidence that the mechanisms in white and African Caribbean populations differ at a crucial point. Thus, the general mechanistic claim that the mechanisms between these two populations are similar is ruled out (indicated by #). This is why instead calcium channel blockers are the recommended anti-hypertensive treatment in African Caribbean populations (Clarke and Russo 2016).

**Example:** *Evaluating dose-response relationships.*

A particular challenge in evaluating the effects of a pharmacological intervention, or effects of an exposure to a chemical agent considers dose-response behaviour. Typically, dose-response is not linear, as metabolic pathways will eventually saturate as the dose increases. It may also be the case that the rate of metabolism and types of metabolites produced vary at specific doses. Normally, one does not have experimental or other data on dose-response at every level of clinical or public health interest. Rather, effects of very low or high doses must be inferred relying on models fitted to whatever data are available. This creates an extrapolation problem—how to establish that the projected responses are accurate, i.e., that the extrapolation from observed data points is reliable. Hypotheses about mechanisms often need to be considered here. For instance, assuming that dose-response is linear, and inferring hypothetical low (respectively, high) dose responses from this assumption implies that the same mechanisms, operating in the same way, are responsible for the response at all or most dose ranges. If, in contrast, measured or estimated responses suggest dose-specific effects (in the form of non-linear dose-response curve), this implies competition between dissimilar metabolic mechanisms.

An example of such an extrapolation problem comes from research on benzene. Recent evidence suggests that benzene is metabolised more rapidly at low exposures, and that low-exposure metabolism favours more hazardous metabolites (Thomas et al. 2014). If true, this implies that different mechanisms operate at low exposures than higher ones. These mechanisms should be such that they are highly sensitive to benzene—i.e., involve a high-affinity enzyme—but are quickly saturated, wherein metabolism switches to other mechanisms as the exposure increases (Rappaport et al. 2009). Estimating very low exposure levels and measuring the response can be methodologically challenging, forcing researchers to engage in extrapolations described above. Mechanistic evidence thus becomes crucial—more direct evidence of the features of enzymatic components of a metabolic mechanism that has high affinity, but gets quickly saturated, is called for. As of now, the question of low-exposure effects of benzene remains open to debate.

## References

- AHRQ (2013). Mechanistic evidence in evidence-based medicine: A conceptual framework. Technical report. Agency for Healthcare Research and Quality. [https://ahrq-ehc-application.s3.amazonaws.com/media/pdf/mechanistic-evidence-framework\\_white-paper.pdf](https://ahrq-ehc-application.s3.amazonaws.com/media/pdf/mechanistic-evidence-framework_white-paper.pdf).
- Aronson, J. (2005). Biomarkers and surrogate endpoints. *British Journal of Clinical Pharmacology*, 59(5), 491–494.

- Chow, S.-C., & Liu, J.-P. (2004). *Design and analysis of clinical trials*. Hoboken, New Jersey: Wiley.
- Clarke, B., & Russo, F. (2016). Causation in medicine. In J. Marcum (Eds.), *Companion to contemporary philosophy of medicine* (pp. 297–322). Bloomsbury.
- Fuhrquist, F., & Saijonmaa, O. (2008). Renin-angiotensin system revisited. *Journal of Internal Medicine*, 264, 224–236.
- Hill, A. B. (1965). The environment and disease: association or causation? *Proceedings of the Royal Society of Medicine*, 58, 295–300.
- Holman, B. (2017). Philosophers on drugs. *Synthese*. <https://doi.org/10.1007/s11229-017-1642-2>.
- Khan, J., & Beevers, D. (2005). Management of hypertension in ethnic minorities. *Heart*, 91(8), 1105–1109.
- Leenen, F. H., Nwachuku, C. E., Black, H. R., Cushman, W. C., Davis, B. R., Simpson, L. M., et al. (2006). Clinical events in high-risk hypertensive patients randomly assigned to calcium channel blocker versus angiotensin-converting enzyme inhibitor in the antihypertensive and lipid-lowering treatment to prevent heart attack trial. *Hypertension*, 48(3), 374–384.
- Montgomery, D. (2009). *Design and analysis of experiments*. Hoboken, New Jersey: Wiley.
- Piller, L. B., Ford, C. E., Davis, B. R., Nwachuku, C., Black, H. R., Oparil, S., et al. (2006). Incidence and predictors of angioedema in elderly hypertensive patients at high risk for cardiovascular disease: A report from the antihypertensive and lipid-lowering treatment to prevent heart attack trial (ALLHAT). *The Journal of Clinical Hypertension*, 8(9), 649–658.
- Rappaport, S. M., Kim, S., Lan, Q., Vermeulen, R., Waidyanatha, S., Zhang, L., et al. (2009). Evidence that humans metabolize benzene via two pathways. *Environmental Health Perspectives*, 117(6), 946.
- Thomas, R., Hubbard, A. E., McHale, C. M., Zhang, L., Rappaport, S. M., Lan, Q., et al. (2014). Characterization of changes in gene expression and biochemical pathways at low levels of benzene exposure. *PloS one*, 9(5), e91828.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



# Chapter 7

## Using Evidence of Mechanisms to Evaluate Efficacy and External Validity



**Abstract** Previous chapters in Part III develop accounts of how to gather and evaluate evidence of claims about mechanisms. This chapter explains how this evaluation can be combined with an evaluation of evidence for relevant correlations in order to produce an overall evaluation of a causal claim. The procedure is broken down to address efficacy, external validity, and then the overall presentation of the claim.

In this chapter, we move from claims about mechanisms to causal claims, i.e., claims of efficacy and external validity. As we have seen in Chap. 6, in order to establish efficacy, one needs to establish both the claim that there is a correlation between putative effect and putative cause and the claim that there is a mechanism connecting the putative effect and cause that can account for the size of the observed correlation. Sect. 7.1 explains how these two types of evidence can be combined to evaluate the status of an efficacy claim. For purposes of clinical or public health decision making one often wants to make inferences about effectiveness, i.e., about causality in target populations other than the study population. Besides evidence directly about the target population, evidence of mechanistic similarity between the target populations and study populations for which efficacy has already been evaluated may be relevant to the status of the causal claim in the target population. We deal with this question of external validity in Sect. 7.2.

### 7.1 Efficacy

Here we address the question of how to combine evaluations of a general mechanistic claim and a correlation claim in order to evaluate a claim of effectiveness.

**General mechanistic claim.** We have seen (in Chap. 6) that the status of the claim that there is a mechanism connecting putative cause and effect is assessed along two different dimensions:

1. Is clinical study evidence strong enough to make it plausible that there is a mechanism that can account for the size of the correlation?
2. Is there a specific mechanism hypothesis and is the existence of the crucial features of *that* mechanism hypothesis established?

**Correlation claim.** The correlation claim is the claim that there is a [correlation](#) between the putative cause and effect, conditional on plausible confounders. Note that mechanistic evidence and results from previous clinical studies may rule in some variables as plausible confounders. Mechanistic evidence may also speak to the question whether a certain clinical study is well-conducted and properly controlled for these confounding variables. Given that one has settled on both a set of potential confounders and an assessment of the quality of the design of the relevant studies, deciding whether the putative cause and effect are correlated is a purely statistical question. A meta-analysis, for instance, of relevant studies yields an estimate for the size of the correlation and corresponding confidence interval and  $p$ -value. The status of the correlation claim then depends on the width of the confidence interval, the size of the  $p$ -value, and the heterogeneity of the studies evaluated. A low  $p$ -value may, for instance, lead to a high status of the correlation claim.

**Efficacy claim.** To obtain the status of an efficacy claim, we combine the status of the corresponding general mechanistic claim with the status of the corresponding correlation claim. Efficacy is established just when it is established that there is a correlation and that there is some mechanism which can account for this correlation (Russo and Williamson 2007; Illari 2011; Clarke et al. 2013, 2014). More generally, the status of the causal claim can be taken to be the minimum of two statuses: the status of the correlation claim and the status of the general mechanistic claim:

**Status of an efficacy claim.** The status of the claim  $A$  is a cause of  $B$  is the minimum of:

1. the status of the claim that  $A$  and  $B$  are appropriately correlated, and
2. the status of the claim that there is an appropriate mechanism linking  $A$  and  $B$  that can account for this correlation.

Hence, a causal claim cannot have a higher status than both the correlation claim and the general mechanistic claim (see discussions in (Russo and Williamson 2007, 2011, 2012; Russo 2011; Clarke and Russo 2016, 2017)). To give an example, efficacy is provisionally established if the existence of a correlation is established or provisionally established and the existence of a mechanism that can account for the correlation is provisionally established. Equally, efficacy is provisionally ruled out if a correlation is provisionally ruled out and if the existence of a mechanism that can account for the correlation is provisionally ruled out or of higher status.

Before turning to external validity, we discuss a potential source of confusion:

**Digression: reinforced concrete.** In the framework set out above, there are two separate distinctions in play. First, there is the distinction between evidence of correlation and evidence of mechanisms (Illari 2011). This distinction is core to the approach

taken in this handbook: the claim that *A* is a cause of *B* is evaluated according to how strongly evidence of correlation supports the claim that *A* and *B* are appropriately correlated, and how strongly evidence of mechanisms supports the general mechanistic claim that there is a mechanism linking *A* to *B* that can account for the correlation. Second, there is a distinction between **clinical studies** (which repeatedly measure *A* and *B* together) and **mechanistic studies** (which investigate the details of a putative mechanism linking *A* and *B*). It is important to note that these two distinctions do not align. Both clinical and mechanistic studies can provide evidence of correlation (though clinical studies often provide better evidence of correlation than mechanistic studies). Similarly, both clinical and mechanistic studies can provide evidence of mechanisms (although mechanistic studies often provide better evidence). See Fig. 3.1. Moreover, there are situations in which a causal claim can be established on the basis of clinical studies alone, as explained in Sect. 2.3 and Chap. 6.

Clinical studies and mechanistic studies can be mutually reinforcing. Consider an analogy to reinforced concrete, which is formed by placing steel grids into concrete (Clarke et al. 2014). Concrete has high resistance to compressive stresses but fractures under tension. Steel, however, has high strength in tension. So, if steel is placed in concrete to produce reinforced concrete, we get a composite material where the concrete resists the compression and the steel resists the tension. The combination of two different materials produces a material that is much stronger than either of its components. In the same way, combining clinical studies with mechanistic studies produces much stronger overall evidence of efficacy than would either type of evidence on its own, because they compensate for each other's weaknesses. For instance, clinical studies can rule out masking: masking occurs when one or more counteracting mechanisms cancel out the effect of the mechanism of action. On the other hand, mechanistic studies can rule out confounding.

The following scenarios illustrate the idea of reinforced concrete.

*Scenario 1.* Suppose, for instance, that many well conducted RCTs consistently show a correlation between the putative cause and effect and that bench research provides only very low quality evidence for the general mechanistic claim that there exists a mechanism that can account for the size of the correlation. In this case, it might seem that the correlation is established and the existence of the mechanism is speculative. In which case, efficacy is only speculative. However, this misrepresents the evidence for the general mechanistic claim. It confuses evidence obtained *only* by bench research with total evidence of mechanisms from all sources. Recall from Sect. 6.3 that clinical studies may also yield evidence relevant to the general mechanistic claim that there exists a mechanism—see Joffe (2011) and Williamson (2018, Sect. 2.1). In the above example, the RCTs, when combined with the bench research, can yield a status for the general mechanistic claim that is higher than speculative—an application of the reinforced concrete metaphor. Accordingly, the efficacy claim will have a status higher than speculative.

*Scenario 2.* Suppose low quality clinical studies suggest that there is a correlation. Suppose too that high quality mechanistic studies support key aspects of a specific mechanism hypothesis, but that the possibility of a counteracting mechanism cannot be ruled out. In this case, it is not clear that the proposed mechanism of action can

account for the observed correlation, and the general mechanistic claim will not be established. Subsequently, high quality clinical studies are carried out and determine that the net correlation is indeed positive. These studies provide evidence that any counteracting mechanism fails to totally mask the effect of the mechanism of action. The total body of evidence may now suffice to establish the general mechanistic claim (see Sect. 6.3). In this scenario, clinical studies reinforce mechanistic studies when evaluating the general mechanistic claim.

*Scenario 3.* Suppose certain clinical studies provide low quality evidence of a correlation. One might think that the key concern is confounding, so that when there is high quality evidence of mechanisms that rules out confounding, efficacy is established. However, confounding is not the only problem that arises with low quality evidence of correlation. There is also the problem that the observed correlation may not correspond to a correlation in the underlying data-generating probability distribution. In order to establish efficacy, one needs to establish that there is a genuine correlation in the underlying distribution. Hence, without high quality evidence of correlation, efficacy cannot be established.

*Scenario 4.* Suppose that initially, certain clinical studies provide low quality evidence of a correlation. Suppose that in this case, it is clear that the studies identify a genuine correlation conditional on certain potential confounders, but that not all plausible confounders have been controlled for. The key concern here, then, is confounding. For instance, there might be a large number of epidemiological studies all showing a correlation between putative cause and effect, but where each study fails to control for some particular variable which may be a confounding variable. Now, if there is also high quality evidence of mechanisms that rules out this variable as a confounder, efficacy is established. In this case, the mechanistic studies boost the status of the correlation claim, to established. In this case, then, the overall status is established.

## 7.2 External Validity

When mechanisms within a study population and the target population are sufficiently similar, one can extrapolate an efficacy claim from the study population to the target population. In this section, we show how to combine evidence of efficacy obtained directly on the target population with evidence obtained by extrapolation from a study population.

Three assessments feed into the evaluation of effectiveness:

1. *Efficacy in the target population.* Although studies performed directly on the target population will normally be less conclusive than those performed on the study population, they can form the basis of a preliminary evaluation of efficacy in the target population. The preliminary status of the causal claim can be determined as set out in Sect. 7.1.

**Table 7.1** Determining the status of the causal claim in the target population given the status of the causal claim in the study population, the status of the claim that the mechanisms of action in study and target are similar, and the status of the causal claim in the target population on the basis only of studies carried out on the target population

		Causation in <b>study</b> population + <b>similarity</b> of mechanism in target and study				
		Established + established	Provisionally established + established; or Established + provisionally established	Other combinations	Provisionally ruled out + established; or Ruled out + provisionally established	Ruled out + established
Causation in <b>target</b> from target studies	Established	Established	Established	Established	Established	Provisionally established
	Provisionally established	Established	Provisionally established	Provisionally established	Provisionally established	Arguable
	Arguable	Established	Provisionally established	Arguable	Speculative	Speculative
	Speculative	Established	Arguable	Speculative	Arguably false	Ruled out
	Arguably false	Speculative	Speculative	Arguably false	Provisionally ruled out	Ruled out
	Provisionally ruled out	Arguably false	Provisionally ruled out	Provisionally ruled out	Provisionally ruled out	Ruled out
	Ruled out	Provisionally ruled out	Ruled out	Ruled out	Ruled out	Ruled out

2. *Efficacy in the study population.* The status of efficacy in the study population can also be determined by considering the procedure of Sect. 7.1.
3. *Similarity of mechanisms in the study and target populations.* The status of the general mechanistic claim relevant to external validity (i.e., the claim that the mechanisms of action are sufficiently similar in study and target) can be determined as indicated in Sect. 6.3.

To obtain a final status for efficacy in the target, one can combine the preliminary status in the target population with the status of efficacy in a study population, provided that study and target population share similar mechanisms of action. The status of the causal claim about the target population may be increased (respectively, decreased) by observing that efficacy does (respectively, does not) hold in a study population that is similar to the target population. In this case, causal claims are extrapolated from the study population to the target population.

Table 7.1 shows how the **status** of the causal claim in the target population can be determined from the above three assessments. To change the preliminary status of an efficacy claim given by studies directly on the target population, all evidence of causation in the study population and of similarity of mechanisms needs to be of at least moderate quality, and one or other needs to be high quality. Other quality levels do not change the initial status.

Some remarks help to explain the table and relate it to other approaches that address external validity.

1. If studies on the target population would on their own establish causality in the target population, this is strong, but not infallible, evidence for causation in the target. If there is a study population for which similarity to the target has been established but causation has been ruled out in the study population, then causation in the target population is downgraded to *provisionally established*. (Note that this situation is not covered by the protocol for evaluating external validity advocated by the International Agency for Research on Cancer (IARC); see Sect. 8.1 for further discussion of this point.)
2. Changing the preliminary status of a causal claim obtained from evidence gathered on the target population is more common when that evidence is of lower quality. For instance, a *provisionally established* status may be changed to *established* only in case of established efficacy in the study and established similarity between study and target. The status *arguable*, however, may be changed in case of established efficacy in the study and provisionally established similarity.
3. The GRADE working group also evaluates whether evidence from a study population can be used to draw inferences about the target population. In particular, the GRADE working group considers the case where no evidence directly obtained on the target population is available:

In general, one should not rate down for population differences unless one has compelling reason to think that the biology in the population of interest is so different from that of the population tested that the magnitude of effect will differ substantially. Most often, this will not be the case. [...] The above discussion refers to different human populations, but sometimes the only evidence will be from animal studies, such as rats or primates. In general, we would rate such evidence down two levels for indirectness (Guyatt et al. 2011, pp. 1304–1305)

Hence, the GRADE working group takes similarity of mechanisms to be *established* by default when study and target populations are both human populations. This is problematic because it sets the standard of evidence required for extrapolation too low. In the case of animal studies, one can interpret the default assumption of the GRADE working group as being that the causal claim is *arguable* solely on the basis of causation in animals having been established. Again this is problematic. In our approach, in the absence of evidence of similarity of mechanisms, efficacy in the study population cannot be extrapolated to the target. Hence, even if many high quality RCTs in animals establish efficacy in animals, in the absence of evidence of similarity, nothing can be concluded about efficacy in humans. There is thus a sense in which the approach presented here is more cautious than the GRADE approach to external validity.

4. Causation can be established or ruled out even where no clinical studies on the target are available. This is the case when causation has been established in a study population for which it has been established that it is mechanistically similar to the target population. (This case is captured by the fourth row of Table 7.1, where causation in the target is *speculative*.)
5. Note that, by similarity of mechanisms we mean that any mechanisms in the target population which counteract this mechanism do not mask the effect of the mechanism of action to such an extent that a net correlation in the target

population could not be explained mechanistically (see Sect. 2.3). Consequently, with a mechanism established and some counteracting mechanisms established in the study, a small correlation may be good evidence for causation in the target even if it is not the case that the whole mechanistic structure is similar. After all, this counteracting mechanism would only make the existent correlation smaller in the study than in the target.

### 7.3 Presenting the Status of a Causal Claim

In presenting the status of a causal claim the following questions need to be addressed, and the status of the causal claim presented after the evaluation of evidence.

**Presenting the status of the efficacy claim.** The following questions should be addressed:

1. What is the population to which the status applies?
2. What is the intervention or exposure level?
3. What is the outcome and how is it measured?
4. What is the status of the correlation claim? How is this status obtained?
5. What is the status of the general mechanistic claim? How is this status obtained? (See Chap. 6.)
6. What is the status of the efficacy claim?

The following box considers the case where efficacy is extrapolated from one to another population

**Presenting the status of the effectiveness claim.** The following questions should be addressed:

1. What is the target population to which the status applies?
2. What is the intervention or exposure level in the target?
3. What is the outcome and how is it measured in the target?
4. What is the study population?
5. What is the intervention or exposure level in the study?
6. What is the outcome and how is it measured in the study?
7. What is the status in the study? How is this status obtained?
8. What is the status in the target obtained by evidence directly of the target? How is this status obtained?
9. What is the status of the general mechanistic claim, i.e. that target and study are similar? (See Chap. 6.)
10. What is the overall status of the effectiveness claim?

Standard evidence appraisal systems can be extended to take these considerations into account. For an example of how to incorporate certain aspects of this procedure into a GRADE-style evidence profile, see Sect. 4.6.

## References

- Clarke, B., & Russo, F. (2016). Causation in medicine. In J. Marcum (Ed.), *Companion to contemporary philosophy of medicine* (pp. 297–322). Bloomsbury.
- Clarke, B., & Russo, F. (2017). Mechanisms and biomedicine. In S. Glennan & P. Illari (Eds.), *The Routledge handbook of mechanisms and mechanical philosophy, Chap. 24* (pp. 319–331). London: Routledge.
- Clarke, B., Gillies, D., Illari, P., Russo, F., & Williamson, J. (2013). The evidence that evidence-based medicine omits. *Preventive Medicine, 57*(6), 745–747.
- Clarke, B., Gillies, D., Illari, P., Russo, F., & Williamson, J. (2014). Mechanisms and the evidence hierarchy. *Topoi, 33*(2), 339–360.
- Guyatt, G., Oxman, A. D., Akl, E. A., Kunz, R., Vist, G., Brozek, J., et al. (2011). Grade guidelines: 1. Introduction grade evidence profiles and summary of findings tables. *Journal of Clinical Epidemiology, 64*(4), 383–394.
- Illari, P. M. (2011). Mechanistic evidence: Disambiguating the Russo-Williamson thesis. *International Studies in the Philosophy of Science, 25*(2), 139–157.
- Joffe, M. (2011). The gap between evidence discovery and actual causal relationships. *Preventive Medicine, 53*(4–5), 246–249.
- Russo, F. (2011). Causal webs in epidemiology. *Paradigmi, XXXIX*(1), 67–98.
- Russo, F., & Williamson, J. (2007). Interpreting causality in the health sciences. *International Studies in the Philosophy of Science, 21*(2), 157–170.
- Russo, F., & Williamson, J. (2011). Epistemic causality and evidence-based medicine. *History and Philosophy of the Life Sciences, 33*(4), 563–582.
- Russo, F., & Williamson, J. (2012). EnviroGenomarkers: The interplay between mechanisms and difference making in establishing causal claims. *Medicine Studies: International Journal for the History, Philosophy and Ethics of Medicine & Allied Sciences, 3*, 249–262.
- Williamson, J. (2018). Establishing causal claims in medicine. *International Studies in the Philosophy of Science*, in press. <http://blogs.kent.ac.uk/jonw/>.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



**Part IV**  
**Particular Applications**

## Chapter 8

# Assessing Exposures



**Abstract** An important problem in causal inference in medicine involves establishing causal relationships between environmental exposures and negative health outcomes. It is typically not possible to use RCTs to solve this problem, for ethical reasons. The approach outlined in this book is compared to two other prominent approaches: the procedures of the International Agency for Research on Cancer (IARC), and SYRINA, a framework for detecting exposures that affect the endocrine system.

An important problem in causal inference in medicine involves establishing causal relationships between environmental exposures and negative health outcomes (Hill 1965). Experimental studies, e.g., randomized controlled trials, tend to provide relatively strong evidence for causal claims. However, when assessing exposures it is typically not possible to carry out such trials in human populations, because this would involve unethically intervening to expose individuals to factors that are suspected to have deleterious health effects. The only available epidemiological studies are observational. As a result, it is difficult to obtain epidemiological data that are sufficient to establish causality.

This problem occurs, for instance, when assessing whether an environmental exposure is carcinogenic in humans. In such cases, different types of evidence are required. For example, the International Agency for Research on Cancer (IARC) attempts to determine whether particular exposures cause cancer in humans by looking at a variety of different types of evidence, namely, epidemiological studies, studies in experimental animals, and mechanistic and other relevant data (IARC 2015). The problem also occurs in assessing whether an exposure is an endocrine disruptor. In this context, Vandenberg et al. (2016) introduced SYRINA, a framework for the systematic review and integrated assessment of exposures. In this chapter, we compare the approach to assessing exposures given in this book with these other prominent approaches. First compare our approach to external validity to the approach endorsed by IARC, with reference to the example of establishing carcinogenicity of

benzo[a]pyrene—a compound that IARC recently evaluated and decided to upgrade from probable human carcinogen to human carcinogen largely based on just the mechanistic evidence and evidence from cancer bioassays. We then compare our approach to SYRINA, a framework for detecting exposures that affect the endocrine system (Sect. 8.2).

## 8.1 Comparison to IARC

Here we compare our approach to external validity to that of the International Agency for Research on Cancer (IARC). A note on terminology here. IARC use the term *generalizability*, as well as external validity, and for the purpose of this discussion we will regard them as synonymous. First, consider an example:

**Example:** *Carcinogenicity of benzo[a]pyrene.*

Benzo[a]pyrene is a polycyclic aromatic hydrocarbon (PAH) that is formed during incomplete combustion of organic material. Benzo[a]pyrene and other PAHs are an important industrial pollutant in soil, water, air, and sediments. They are also found in high concentrations in tobacco smoke, and in some pharmaceutical products. Human exposure occurs mainly through industrial and environmental exposure (IARC 2009). IARC has evaluated benzo[a]pyrene in four monographs, and it is currently classified as Group 1, carcinogenic to humans (IARC 2015).

In the most recent evaluation, epidemiological data were not available to the IARC working group. The working group therefore made its decision to classify benzo[a]pyrene as carcinogenic to humans based on mechanistic evidence and evidence from experimental animals. This makes the case of benzo[a]pyrene especially interesting for our purposes, as according to the procedure outlined above in Sect. 7.2, the correlation between benzo[a]pyrene and cancer required to establish the causal claim in humans would have to be inferred from observed outcomes in the experimental animals together with the mechanistic data.

On the approach of this book, first one formulates the causal claim under scrutiny: here, ‘benzo[a]pyrene causes cancer in humans’. In the context of IARC, this is to be taken as a qualitative claim—IARC identifies cancer hazards, and the exact size of the effect by which exposure increases cancer risk does not play a role in determining carcinogenicity. We should note though that a qualitative understanding of effect size does play a role in determining carcinogenicity. The IARC process is explicitly based on the causal indicators set out by Hill (1965), as we discuss above.

Next, one should assess—according to a suitable framework—the evidence for a correlation between the exposure and its effect, and articulate any hypo-

thetical mechanisms that would account for the correlation. Note that IARC use their own framework for assessing correlations (IARC 2015). A GRADE-like framework would be potentially useful in this context too—assuming that suitable modifications can be made to allow for differences in the understanding of bias in evidence that are appropriate for this change in purpose.

The evidence for the relevant mechanisms should then be graded according to the procedures described in Chap. 6. In the latest IARC monograph on benzo[a]pyrene, all the evidence of a correlation between the exposure and cancer came from studies on experimental animals—no epidemiological data were evaluated. The correlation between exposure and cancer in humans must thus be inferred via extrapolation from corresponding data in the experimental animals. This is based on assessing the evidence for correlation in the experimental animals, and assessment of similarities of the underlying mechanisms. The IARC monograph reports evidence of cancer outcomes upon exposure to benzo[a]pyrene in experimental animals. This was judged to be of high quality, both in terms of the validity of the research within species of experimental animals, and in terms of the additional corroboration gained by these results being robust across eight species of experimental animals (IARC 2009, 112–131). In addition, evidence is presented and evaluated for two main types of mechanism by which benzo[a]pyrene causes DNA adducts to form at known cancer hotspots: in one of these a metabolite of benzo[a]pyrene binding the DNA molecule, and the other an oxidized form of benzo[a]pyrene. In addition, similar activity of benzo[a]pyrene is reported to be shown in *in vitro* studies on human cell lines (IARC 2009, 131–137).

IARC considered there to be sufficient evidence for carcinogenicity in the experimental animals, *i.e.*, the causal claim about the experimental animals was established. IARC's current practice is to make some evaluations about possible mechanisms of carcinogenesis using a set of key characteristics shown by carcinogens (Smith et al. 2016). This is broadly compatible with the approach of this book, as there is high quality evidence of both correlation and underlying mechanisms in the experimental animals. This alone would not suffice to transfer the same claim to humans (nor does the IARC approach consider this). However, strong evidence of similar mechanisms operating in the experimental animals and humans, and the robustness of the experimental animal results across many species, warrants a mechanism-based extrapolation of the causal claim from the experimental animals to humans (Wilde and Parkkinen 2017). This, together with the mechanistic evidence directly on humans, such as evidence of formation of DNA adducts, is what, on the approach presented here, warrants establishing a causal conclusion about humans. In mechanism-based extrapolation, one compares the mechanisms responsible for an outcome in the target—of which a conclusion about causality is to be made—and in the study—about which direct evidence of causality is available—and looks for differences that might lead to differences in the outcome of interest between the study and the target. Here the outcome of interest is the development of tumours or the appearance of various cancer biomarkers upon exposure to

benzo[a]pyrene. A dependence between these outcomes and benzo[a]pyrene has been robustly demonstrated in the experimental animals. The relevant mechanisms are the pathways by which benzo[a]pyrene causes DNA adducts that can trigger tumorigenesis, that would explain the dependence. For these, there is evidence from cultured human cell lines, as well as the experimental animals, demonstrating strong similarities, and no differences that would indicate that benzo[a]pyrene does not cause cancer in humans. In addition, there is concordant evidence of the outcomes in several species of experimental animal, lending further credibility to the assumption that the carcinogenicity of benzo[a]pyrene is not dependent on idiosyncratic features of any particular species. These considerations, taken together, suffice to establish the carcinogenicity of benzo[a]pyrene in humans.

While the approach of this book would yield the same conclusion as IARC's, it should be noted that the procedures differ at certain points. IARC does not formally endorse extrapolation from experimental animals. Note though that this does not preclude altogether judgements about possible carcinogens where no human research is available, as in cases where only animal studies are available substances may be classified by IARC as belonging to *Group 2B: The agent is possibly carcinogenic to humans*. Nor does IARC formally endorse robustness of evidence as grounds for upgrading a classification, but allows for upgrading (or downgrading) a classification of carcinogenicity on the basis of mechanistic evidence alone. On the approach of this book, one may appeal to the aforementioned considerations, and one needs in addition to establish correlation in humans (by direct observation or extrapolation), before any claim about causality can be considered established.

Having considered an example, we now compare the general approach of this book to external validity to that of IARC. IARC's approach is summarized in Fig. 8.1.

The categories of IARC roughly correspond to those presented here, as follows. IARC have a ranking for overall carcinogenicity:

- Group 1* : Established
- Group 2a* : Provisionally established
- Group 2b* : Arguably true
- Group 3* : Speculative
- Group 4* : Ruled out

IARC also has a separate ranking of evidence of carcinogenicity in humans and animals:

- Sufficient* : Established
- Limited* : Provisionally established
- Inadequate* : Arguable or speculative
- Evidence Suggesting Lack of Carcinogenicity (ESLC)* : Ruled out

		EVIDENCE IN EXPERIMENTAL ANIMALS			
		<i>Sufficient</i>	<i>Limited</i>	<i>Inadequate</i>	<i>ESLC</i>
<i>Sufficient</i>		Group 1			
<i>Limited</i>		↑1 strong evidence in exposed humans Group 2A	↑2A belongs to a mechanistic class where other members are classified in Groups 1 or 2A Group 2B (exceptionally, Group 2A)		
EVIDENCE IN HUMANS	<i>Inadequate</i>	↑1 strong evidence in exposed humans ↑2A strong evidence ... mechanism also operates in humans Group 2B	↑2A belongs to a mechanistic class ↑2B with supporting evidence from mechanistic and other relevant data Group 3	↑2A belongs to a mechanistic class ↑2B with strong evidence from mechanistic and other relevant data Group 3	Group 3 ↓4 consistently and strongly supported by a broad range of mechanistic and other relevant data
	<i>ESLC</i>	↓3 strong evidence ... mechanism does not operate in humans	Group 3		

Fig. 8.1 IARC’s approach to classifying potential carcinogens (<http://monographs.iarc.fr/ENG/Publications/Evaluations.pdf>)

In addition, IARC has a separate ranking of evidence of mechanisms:

- Strong* : Established
- Moderate* : Provisionally established
- Weak* : Arguable or speculative

What is being assessed by these three categories is a general mechanistic claim: e.g., the existence of a mechanism of action in animals; or the similarity of mechanism of action in humans to that in animals; or the existence of a mechanism of action in humans.

The approach of this book is simpler than that of IARC in one respect: a single scale from established to ruled out, rather than three different categorisations. On the other hand, the scale adopted in this book involves more categories.

In order to compare the approach of this book with that of IARC, consider two tables that illustrate the approach that this book takes with respect to external validity. First, Table 8.1 assumes that causality in the study has been established and charts similarity of mechanisms in the study and target populations against causation in the target population on the basis of evidence obtained on the target population. A second table, Table 8.2, assumes that similarity of mechanism is established and charts causation in the study population against causation in the target population on the basis of evidence obtained in the target population.

**Table 8.1** Determining the status of the causal claim from similarity of mechanisms in the study and target populations and causation in the target population on the basis of evidence obtained on the target population. It is assumed here that causality in the study population has been established

		Similarity of mechanism in target and study				
		Established	Provisionally established	Arguable	Speculative	Ruled out
Causation from target studies	Established	Established	Established	Established	Established	Established
	Provisionally established	Established	Provisionally established	Provisionally established	Provisionally established	Provisionally established
	Arguable	Established	Provisionally established	Arguable	Arguable	Arguable
	Speculative	Established	Arguable	Speculative	Speculative	Speculative
	Ruled out	Provisionally ruled out	Ruled out	Ruled out	Ruled out	Ruled out

**Table 8.2** Determining the status of the causal claim from causation in the study population and causation in the target population on the basis of evidence obtained on the target population. It is assumed here that similarity of mechanism has been established

		Causation in the study population				
		Established	Provisionally established	Arguable	Speculative	Ruled out
Causation from target studies	Established	Established	Established	Established	Established	Provisionally established
	Provisionally established	Established	Provisionally established	Provisionally established	Provisionally established	Arguable
	Arguable	Established	Provisionally established	Arguable	Arguable	Speculative
	Speculative	Established	Arguable	Speculative	Speculative	Ruled out
	Ruled out	Provisionally ruled out	Ruled out	Ruled out	Ruled out	Ruled out

There is a broad agreement between the approach presented here and that of IARC. As with the approach advocated here, IARC employs evidence of mechanisms to draw conclusions about causation at two places: to evaluate efficacy in humans on basis of evidence directly in humans and to ensure that causal claims in specific animal populations can be extrapolated to humans. For the first task, IARC employs the Hill indicators without assessing mechanistic studies in a systematic way. It is only in assessing external validity that IARC explicitly evaluates studies that investigate the details of the mechanism of action.

The approach presented here is more explicit with respect to where and what evidence of mechanisms should be used. Firstly, this book recommends explicitly evaluating mechanistic studies when evaluating evidence obtained directly in humans. After all, evaluating both whether there exists a mechanism and whether

there exists a correlation is necessary for evaluating the evidence obtained directly in humans (Sect. 7.1). The Hill indicators can only be seen as a first approximation to the comprehensive assessment of mechanistic evidence needed to establish efficacy in humans. What is more, these indicators tend to obfuscate, rather than clarify, distinctions between evidence pertinent to the correlational claim and evidence pertinent to the general mechanistic hypothesis (Chap. 6).

Secondly, this book separates the overall evaluation of causality and the evaluation of evidence directly obtained in humans. The overall evaluation is obtained by aggregating the evidence directly obtained in humans and the evidence in animals (Sect. 7.2). For instance, it might be that, initially, some causal claim is established in humans by considering studies that purely involve humans, but that, subsequently, studies of a variety of animal species that are mechanistically similar to humans rule out causation in those species. These further studies would surely cast enough doubt on causation in humans so that the causal claim can no longer be considered established. However, by identifying the overall evaluation with the evaluation of evidence directly obtained in humans when the evidence obtained on humans is sufficient (see the top row of the IARC table, Fig. 8.1), IARC assigns Group 1 in this case (the top right-hand corner of the IARC table). The procedure set out in this book would assign status *established* to the causal claim on basis of just the evidence directly obtained in humans, but it would assign overall status *provisionally established* on the basis of all the evidence, animal as well as human (see the top-right corner of Table 8.2). This classification is perhaps more appropriate.

## 8.2 Comparison to SYRINA

SYRINA is a framework that was put forward to evaluate the strength of evidence that a certain exposure is an endocrine disruptor (Vandenberg et al. 2016). This approach first evaluates the evidence for an association between chemical exposure and (adverse) effect. Second, this approach evaluates the evidence for an association between the chemical and endocrine disrupting activity. Third, the evidence for an association with an (adverse) effect and for an endocrine disrupting activity are combined to obtain an overall assessment of endocrine disruption.

SYRINA combines quality of evidence ratings from different streams of evidence in all three steps. As with our approach, the quality level of the causal claim is the minimum of the quality of the different evidence streams. Figure 8.2 gives the relevant SYRINA table for an association between chemical exposure and (adverse) effect.

The resulting initial rating can be upgraded by one level if there is high confidence in the evidence from *in silico* and *in vitro* studies.

In the next step, the endocrine disrupting activity of the exposure is evaluated by combining different evidence streams. This *in vivo* and *in vitro* evidence is combined. Figure 8.3 gives the relevant SYRINA table.

**Fig. 8.2** SYRINA table for an association between chemical exposure and (adverse) effect

Human/Wildlife (observational)	High	Strong	Strong	Strong	Strong
	Medium	Moderate	Moderate	Moderate	Strong
	Low	Weak	Weak	Moderate	Strong
	Absent	No data	Weak	Moderate	Strong
	Absent	Low	Medium	High	
	Experimental <i>in vivo</i>				

**Fig. 8.3** SYRINA table for combining *in vivo* and *in vitro* evidence

<i>In vitro</i>	High	Strong	Strong	Strong	Strong
	Medium	Moderate	Moderate	Moderate	Strong
	Low	Weak	Weak	Moderate	Strong
	Absent	No data	Weak	Moderate	Strong
	Absent	Low	Medium	High	
	Experimental <i>in vivo</i>				

Finally, the quality levels for the association with adverse health outcomes and for the endocrine activity are combined according to the table in Fig. 8.4.

In relatively unusual cases the resulting quality level can be upgraded or downgraded by considerations given to the plausibility of the link of disrupting endocrine disrupting activity and outcome.

Let us consider some points of comparison between SYRINA and the approach of this book. First, this book formulates explicit methods for evaluating evidence of mechanisms (Chap. 6). Second, for the evaluation of both endocrine activity and association with adverse health outcomes, SYRINA only combines two kinds of study. When evaluating the plausibility of an association with adverse outcomes, SYRINA combines results from experimental laboratory animals with evidence in humans or wildlife animals. According to the approach presented in this book, application of results from such associations in animals would need to be extrapolated with the help of evidence of mechanisms along the lines of Sect. 7.2. In addition, mechanistic considerations may be relevant when evaluating whether there is an association of the chemical with adverse health outcomes. After all, an observed correlation may be due to confounding. As with IARC, SYRINA makes use of the Hill indicators

<b>Strength of Evidence: Endocrine Disrupting Activity</b>	<b>Strong</b>	Probable EDC	Probable EDC	Known EDC	Known EDC
	<b>Moderate</b>	Possible EDC	Possible EDC	Probable EDC	Known EDC
	<b>Weak</b>	Not classifiable	Not classifiable	Possible EDC	Probable EDC
	<b>No data</b>	Not classifiable	Not classifiable	Possible EDC	Probable EDC
	<b>No data</b>	<b>Weak</b>	<b>Moderate</b>	<b>Strong</b>	
<b>Strength of Evidence: Association Between Exposure and Health Outcome</b>					

**Fig. 8.4** SYRINA table for combining the quality levels for the association and the endocrine activity

for evaluating each stream of evidence and does not explicitly distinguish between evidence of mechanisms and evidence of correlation. Hence, while this book agrees with SYRINA that many evidence streams should be considered when evaluating causal claims, we would emphasise the need for a more systematic integration of evidence of mechanisms and evidence of correlation along the lines of Chaps. 6 and 7.

## References

- IARC (2015). *IARC monographs on the evaluation of carcinogenic risks to humans: Preamble*. Lyon: International Agency for Research on Cancer.
- Hill, A. B. (1965). The environment and disease: Association or causation? *Proceedings of the Royal Society of Medicine*, 58, 295–300.
- IARC (2009). *IARC monographs on the evaluation of carcinogenic risks to humans, vol. 100F: A review of human carcinogens. Chemical Agents and Related Occupations*. International Agency for Research on Cancer, Lyon.
- Smith, M. T., Guyton, K. Z., Gibbons, C. F., Fritz, J. M., Portier, C. J., Rusyn, I., et al. (2016). Key characteristics of carcinogens as a basis for organizing data on mechanisms of carcinogenesis. *Environmental Health Perspectives*, 124, 713–21.
- Vandenberg, L., Agerstrand, M., Beronius, A., Beausoleil, C., Bergman, A., Bero, L., et al. (2016). A proposed framework for the systematic review and integrated assessment (SYRINA) of endocrine disrupting chemicals. *Environmental Health*, 15(74), 1–19.
- Wilde, M., & Parkkinen, V. -P. (2017). Extrapolation and the Russo–Williamson thesis. *Synthese*. <https://doi.org/10.1007/s11229-017-1573-y>.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



# Chapter 9

## Assessing Mechanisms in Public Health



**Abstract** Further considerations need to be borne in mind for evidence appraisal in areas beyond clinical medicine, such as public health. This chapter looks at how public health has treated associations and correlations. Then it examines the importance to public health of mechanisms operating at the group and individual level, concerning social interactions and support, access to socio-sanitary infrastructures, psychological factors, and so on, which have to be explored in the appraisal of public health evidence. Finally, the chapter considers the relationship between biological and social factors, and the difference between mechanisms of disease and mechanisms of prevention.

### 9.1 Introduction

When applying the ideas described in this book to areas other than therapeutic clinical medicine, a number of further considerations need to be borne in mind. The arena beyond clinical medicine where most thinking has been done relating to methods of evidence appraisal is public health (NICE 2012). Public health is concerned with actions, interventions and policies designed to protect the public from hazards, to prevent disease, and to promote good health (Tannahill 1985). In different countries, specific institutions were given the task of developing methods for the assessment of evidence and for the formulation of guidelines in public health. These individual efforts have been brought together into a European initiative, led by the European Centre for Disease Prevention and Control (ECDC). In their 2011 synthesis report, they show how public health should adopt and integrate the methods of evidence-based medicine, specifically the GRADE system, for the assessment of evidence (European Centre for Disease Prevention and Control 2011). In this chapter the focus is on one particular sub-issue, namely mechanisms of causation and, given the concerns of this book, how to deal with mechanisms conceptually and then practically in the appraisal of evidence.

## 9.2 Public Health and Evidence-Based Medicine in the UK

Public health in the UK has been working within the evidence-based paradigm formally since 2000, and much has been learned (Kelly et al. 2010; Kelly and Moore 2012). In 2001 the English Department of Health published its Research and Development Strategy. Amongst other things it made the case for using the principles of evidence based medicine in public health (Department of Health 2001). Organisations such as the Centre for Reviews and Dissemination at the University of York, the Cochrane Collaboration, the Campbell Collaboration, the Health Development Agency and NICE took up the challenge. These organisations have confronted in various ways the methodological, theoretical, practical, epistemological and ontological problems of applying EBM principles to the very broad church of public health. Since then other policy areas have gone in the same direction of taking an evidence based approach. So social care, education and criminal justice, amongst others, have all had agencies created to move these arenas onto an evidence based footing (Paisley et al. 2018).

## 9.3 Statistical Associations and Correlations in Public Health

Statistical associations and correlations have been at the heart of progress in public health for many years. A number of landmark studies show just how important finding statistical associations can be. The investigations by Doll and Hill (1950, 1952, 1964) into the connections between smoking and disease are the original benchmarks. Their initial observations showed that there was an association between exposure to cigarette smoke and carcinoma of the lung (an association which had not hitherto been noticed). This led, in the long run, to public health policies which have reduced the prevalence of cigarette smoking in the population and greatly reduced the number of deaths from lung cancer, and also heart disease, stroke, and various other cancers, which were subsequently found to be associated with exposure to cigarette smoke too.

These pioneering works are often thought to be purely statistical, but in fact Hill was concerned with biological plausibility, and hence mechanisms (Hill 1965). Since the early 1950s when the first statistical observations were made, the biological mechanisms operating in the interaction between the contents of cigarette smoke and the tissues in the lung, as well as the mechanisms relating to the effects on blood circulation, heart functioning, arterial disease, and many other pathologies have been described. Considerations about biological plausibility also led to investigations of the relation between asbestos and mesothelioma (Doll 1955; Newhouse and Thompson 1965). Scientific discoveries relating to these mechanisms continues to the present. The basic mechanisms are well understood in individual human

beings, and public health policy has developed in such a way that smoking in the European Union is now a minority habit and protection from unwanted exposure to cigarette smoke is the norm.

So cigarette smoking was identified as what public health practitioners have come to call a risk factor. In the wake of this great public health success, statistical associations have emerged over the years pointing to risks from other things, notably a lack of physical activity, being overweight and obese, over consumption of alcohol (Sytkowski et al. 1996), certain types of sexual activity (Dougan et al. 2005), ingesting certain non-prescribed drugs (White and Pitts 1998) as well as toxins in the environment, although the dangerous consequences of exposure to certain substances used in industrial processes like asbestos, phosphorus and radium had been known long before the discoveries about smoking (Gochfeld 2005).

There is now a very large and important scientific literature originating in the observation of statistical correlations and subsequently strengthened into causal understandings based on the mechanisms at work in the human body following exposures. Policies designed to protect the public have flowed from this scientific knowledge. New risks regularly appear and currently the role of air pollution and toxins from emissions from vehicles are under scrutiny. This debate mirrors events in the 1950s when the dangers from smog in urban environments caused by the burning of coal led to the Clean Air Act and the phasing out of coal as a primary domestic fuel in the UK (Brimblecombe 2006). In public health there is a long history of bringing together correlations and mechanisms to understand the processes which can cause a number of very common diseases and which potentially offer a platform to take action to mitigate the risks and harms, and, as with the Clean Air Acts of the 1950s and action against tobacco, have been highly effective and successful.

## 9.4 Recurrent Public Health Problems —Non-communicable Disease in the Present

However, notwithstanding the successes with smoking and clean air, deaths from preventable causes which are known and well understood have not gone away. Deaths from non-communicable diseases associated with excess calorie and alcohol consumption and lack of physical activity continue to increase steadily in most countries around the world (Beaglehole et al. 2012). Type 2 diabetes, cardiovascular disease, and certain cancers all have rising prevalence even though the statistical associations between the diseases and the risk factors are well known and the mechanisms operating at the individual level are well understood (though in some diseases better than others).

This is very important as far as appraising evidence of mechanisms is concerned. It is fundamentally important in ethical terms too, because the rising prevalence, while affecting the whole of the population, affects those in poorer and more disadvantaged

circumstances to a far greater extent than the well to do and the privileged (Wilkinson and Marmot 2003). There is a sharp gradient in health inequities that shows a strong correlation between poor health and early death from non-communicable disease and disadvantage. This holds whether disadvantage is measured by income, occupation (or lack of it), housing tenure or educational level or qualifications (Buck and Frosini 2012). The fact is that there are a number of mechanisms which are conceptually and practically distinct from the mechanisms describing the processes of disease causation following exposure to a pathogen or toxin of some kind. Such mechanisms operate at the group and individual level, and concern social interactions and support, access to socio-sanitary infrastructures, psychological factors, etc. It is these mechanisms as well as the biological ones, which have to be explored in the appraisal of public health evidence (Kelly et al. 2014).

## 9.5 The Individual Level and the Population Level

The first thing to note is that mechanisms operate at different levels. In almost all of the investigations referred to above, the mechanisms that have been subject to most scrutiny are those operating at the level of individual human biology. So, after association were found in the population data, the focus shifted to understanding what was actually going on in the human body when it was exposed to cigarette smoke, ethanol, high levels of sugar, asbestos, particulates in the atmosphere and so on. And this approach of course has shown why these exposures are harmful and how they operate on the human biology. These investigations have been extremely successful and we now have plausible biological mechanistic explanations.

But what about the mechanisms operating at the population level? What about the mechanisms that produce the patterning of health between the rich and the poor, between different parts of countries (Graham and Kelly 2004)? In the United Kingdom, for instance, health on average is much worse in Scotland and the North of England than in the South. How can we explain that? What are the mechanisms which explain the fact that, on average, baby boys born in Guildford will live much longer than baby boys born in Shettleston in Glasgow? What are the mechanisms which link poverty to early death? And what are the relationships between the mechanisms going on biologically and in the wider social and physical environment (Kelly et al. 2014)?

With the stunning progress in understanding the biochemistry of disease since the nineteenth century, the tendency has been to focus on mechanisms operating at the biological individual level. As noted above this is usually relatively straightforward, as the biological processes have been well understood in broad terms for decades and the detail is constantly developing as the science progresses. But what are the social and behavioural mechanisms involved? The behavioral mechanisms are also reasonably well described in the psychological literature (see Table 9.1 for some examples). Models and theories explaining why, on average, humans are likely to do this or that, are plentiful (Conner and Norman 2005). However, why when following

**Table 9.1** Behavioural mechanisms

Hazard	Behavioural mechanisms	Disease/morbidity
Exposure to cigarette smoking	Teenagers imitating peers or smoking parents	Lung cancer, cardiovascular disease
Exposure to ethanol, binge drinking	Socializing	Liver disease, certain cancers
Exposure to HPV	Unprotected sex, socializing	Cervical cancer
Exposure to mosquitos	Sanitation, clothing	Zika
Workplace posture	Incorrect posture while sitting or working with a computer, poor lifting practices	Lower back pain
Work overload	Organizational structures, management practices	Anxiety, depression

the same intervention based on the same information about the dangers of smoking, one individual does “this” (say, decides and successfully quits smoking) and one does “that” (doesn’t even think about quitting smoking) is less well understood in a mechanistic sense (Marteau et al. 2015).

However, where the biggest gaps in mechanistic understandings exist, is at the social or population level. The associations between poverty and poor health have been known since at least the middle of the nineteenth century and for probably much longer than that in a non-statistical sense. But how it works mechanistically is much less well defined. From an evidence appraiser’s point of view there is no easy solution to these problems and neither will there be till primary studies examining the mechanisms have been conducted. But it is important nevertheless to ask the questions. And to ask the questions in a way that acknowledges that we do indeed know with a very high degree of certainty that there is a relationship between wealth, education and employment and health, but we do not know with sufficient clarity what the mechanisms are and in such a way as to target interventions and policies in a directed way to be maximally effective (Kriznik et al. 2018).

There have been many attempts around the world to tackle inequalities in health and while overall the health of populations has improved decade on decade, the relative inequities remain a stubborn fact of life (WHO 2008). Although the lack of political will to do something about it has been a major barrier everywhere, one of the other important reasons for failure has been an absence of mechanistic studies at the population level studies and therefore of the ability to know what to do based on mechanistic understandings of the causal pathways involved.

## 9.6 The Biological Level and the Social Level

In recent years, the relationship between the individual biological level and the social level has come under scrutiny as a consequence of developments in biology itself, particularly developments in developmental programming, epigenetics and metabolomics. While each of these topics is different, what they have in common is that they show how the human phenotype is the product as much of its environment physically and socially as it is of its genetic inheritance (Kelly and Kelly 2018). Human (and animal) biology is much more plastic in the face of environmental exposures than had been previously thought. DNA doesn't change, but the way that it is expressed does. The metabolic structure of our bodies reveals a timeline of the various exposures we have been subjected to across the life course. Factors affecting the health of our grandmother when she was pregnant with our own mother may have a fundamental effect on our own health in adulthood. The mechanisms here are now quite well developed (Hanson and Gluckman 2011; Ozanne and Constância 2007) and they show that our health is not just a metabolic response to toxins; it is about a complex social and biological interaction—a relational process or mechanism. These mechanisms are critically mediated by the social worlds that people inhabit.

This science is still developing at a rapid rate and along with it, the understanding of the human genome and the therefore of individual biological differences between humans. It is highly likely that new and better mechanistic models and understandings will emerge including ones incorporating the social factors. The implications for the evidence appraiser at this stage are that the question should be asked—are mechanisms relating to the relationship between biological and social factors being described, used, and articulated? A further important epistemological consideration is the degree to which the approach taken by the researcher is a genuinely a relational one—in other words, one that sees the process as a dynamic and interactive one rather than a deterministic one. This is important because if the new understandings of the plasticity of biology are to be useful in public health, the models need to move away from a reductionist approach and should instead be about elucidating the interactive nature of the process. Again this is a question to be asked by the evidence appraiser: what is the nature of the interaction?

## 9.7 Mechanisms of Disease and Mechanisms of Prevention

There is another question to be asked about the evidence of mechanisms in public health matters and that is about the difference between the causes of disease and the causes of prevention (Kelly and Russo 2018). So far in this chapter we have focused on the important difference between the causes of disease in individuals and the mechanisms involved and the causes of the patterning of disease at population

**Table 9.2** Public health mechanisms for tackling obesity

Generic, population level	Targeted
Food advertising, e.g., max amount per day, recommended amount per day, amount of lipids or carbohydrates contained in food portions	MEND programmes in the UK <a href="http://www.mendfoundation.org/">http://www.mendfoundation.org/</a> , e.g., targeted training for school children about diet and healthy lifestyles, targeted training for parents about psychological risks related to obesity

level and the mechanisms involved in this patterning. We have also discussed the mechanisms involved in the relationships between the two.

But there is another very important distinction to draw out which is especially important in public health. This is the difference between mechanisms causing the disease (either in individuals or in populations) and the mechanisms involved in preventing disease (e.g., Table 9.2). The question simply is this. Does knowing the cause of a disease (an exposure to something which is risky) and knowing that by reducing exposure that disease will be prevented, tell you how to reduce exposure? The short answer is that it doesn't, though many public health policies proceed as if it did. The biology of the aetiology of lung cancer, of liver disease, of type two diabetes and the metabolic syndrome tell you nothing about the mechanisms involved in helping people to stop smoking, to consume less alcohol, to eat fewer calories or take more exercise. Knowledge of the cause tells us what people should do, but it doesn't explain how to do it. The mechanisms involved in smoking and giving up smoking, the mechanisms involved in the practices of eating and drinking (and for that matter, sexual conduct, bad driving, or going jogging) belong to a quite different realm of evidence than microbiology. The relevant evidence is social and psychological. The mechanisms involved are social and psychological and there is a considerable amount of evidence, some of which has been around for a long time, describing both associations and mechanisms—see Becker et al. (1977) and Kelly and Russo (2018). For the most part, however, public health policy (with the very significant and successful exception of smoking) pays scant attention to the social and psychological evidence, mechanistic or otherwise. We suggest that the evidence appraiser begins by asking the question: what evidence is available about the aetiology of the disease? And what evidence about effective preventive measures? The distinction between aetiology and prevention should then guide the appraisal of correlations and of mechanisms. Specifically, are only mechanisms at biological level invoked, or also social mechanisms?

Finally, for both mechanism of disease and mechanism of prevention, the evidence sources will be heterogeneous. The disciplines of psychology, sociology, economics, anthropology, organisational behaviour, political science, history, and the public health sciences all have, and have had, things to say on these matters. Unfortunately, it is not the case that we can simply cheerfully agree that the evidence for these things is heterogeneous so we should just pull it all together, synthesise it and out will come a nice clear set of mechanisms. The reason for this is that each of these disciplines,

and the many sub-disciplines within each of them, operate with a variety of epistemological, methodological and ontological assumptions about the nature of human life and its place in the world. Sometimes these veer toward highly individualistic accounts sometimes to more socially oriented accounts. So the task is not to try to adjudicate, but to acknowledge the differences, to articulate them (even if the researchers don't themselves do that), and to consider the degree to which the different positions really matter in terms of the substantive problem (Kelly 2017). Intriguingly, all these disciplines are dealing with the same basic concern—humans in the physical and social world and what is going on in their heads as they go about their business. They each construct ways of seeing and describing the same phenomena differently and in ways that sometimes defy any kind of commensurability. However as long as the appraiser keeps in mind that the basic thing under consideration is the same, and there are just lots of different ways of looking at the phenomena, then the task is not an impossible one. But as ever the first step is to ask the appropriate question, to describe what is there in terms of evidence and to determine to what extent this allows us to understand the mechanisms with clarity.

Here are some simple questions that one can ask in order to structure the search for relevant mechanistic studies, in the context of public health interventions:

**Checklist of questions:**

1. What disease is the intervention targeting? Infectious or non-communicable?
2. What biological mechanisms are known?
3. What socio-economic or psychological mechanisms are known scientifically?
4. How can behavioural mechanisms reduce exposure?
5. Why might public health interventions targeting the pathogens fail?
6. What is the public perception of the disease in terms of risk, seriousness and personal vulnerability?
7. What mechanisms come into play as a population or different segments in the population react to an intervention?
8. Are there sub-groups within the population that should be specifically targeted? How can they be reached and what specific mechanisms might come into play?

Users interested in carrying out structured searches for relevant mechanistic studies should refer to the Public Health and Social Care tool in Sect. 4.7.

## References

- Beaglehole, R., Bonita, R., Horton, R., Ezzati, M., Bhala, N., Amuyunzu-Nyamongo, M., et al. (2012). Measuring progress on NCDs: One goal and five targets. *The Lancet*, 380(9850), 1283–1285.
- Becker, M. H., Haefner, D. P., Kasl, S. V., Kirscht, J. P., Maiman, L. A., & Rosenstock, I. M. (1977). Selected psychosocial models and correlates of individual health-related behaviors. *Medical Care*, 15(5), 27–46.
- Brimblecombe, P. (2006). The clean air act after 50 years. *Weather*, 61(11), 311–314.
- Buck, D., & Frosini, F. (2012). *Clustering of unhealthy behaviours over time* (pp. 1–24). London: The Kings Fund.
- Conner, M., & Norman, P. (2005). *Predicting health behaviour*. UK: McGraw-Hill Education.
- Department of Health (2001). A research and development strategy for public health. UK National Archives. <http://webarchive.nationalarchives.gov.uk/20121206164331/>, [http://www.dh.gov.uk/prod\\_consum\\_dh/groups/dh\\_digitalassets/@dh/@en/documents/digitalasset/dh\\_4020946.pdf](http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@en/documents/digitalasset/dh_4020946.pdf).
- Doll, R., & Hill, A. B. (1950). Smoking and carcinoma of the lung. *British Medical Journal*, 2(4682), 739.
- Doll, R. (1955). Mortality from lung cancer in asbestos workers. *British Journal of Industrial Medicine*, 12, 81–86.
- Doll, R., & Hill, A. B. (1952). Study of the aetiology of carcinoma of the lung. *British Medical Journal*, 2(4797), 1271.
- Doll, R., & Hill, A. B. (1964). Mortality in relation to smoking: Ten years' observations of British doctors. *British Medical Journal*, 1(5395), 1399.
- Dougan, S., Gilbert, V. L., Sinka, K., & Evans, B. G. (2005). HIV infections acquired through heterosexual intercourse in the United Kingdom: Findings from national surveillance. *British Medical Journal*, 330(7503), 1303–1304.
- European Centre for Disease Prevention and Control. (2011). How to assess the best available evidence when time is limited and there is lack of sound evidence. European Centre for Disease Prevention and Control, Stockholm: Technical report.
- Gochfeld, M. (2005). Chronologic history of occupational medicine. *Journal of Occupational and Environmental Medicine*, 47(2), 96–114.
- Graham, H., & Kelly, M. P. (2004). *Health inequalities: Concepts, frameworks and policy*. NHS Health Development Agency. Briefing Paper, [https://www.researchgate.net/publication/242494475\\_Health\\_Inequalities\\_Concepts\\_Frameworks\\_and\\_Policy](https://www.researchgate.net/publication/242494475_Health_Inequalities_Concepts_Frameworks_and_Policy).
- Hanson, M., & Gluckman, P. (2011). Developmental origins of noncommunicable disease: Population and public health implications. *The American Journal of Clinical Nutrition*, 94(6 Suppl), 1754S–1758S.
- Hill, A. B. (1965). The environment and disease: Association or causation? *Proceedings of the Royal Society of Medicine*, 58, 295–300.
- Kelly, M. (2017). On epistemological and ontological incommensurability in modelling behaviour change. In S. Christmas, S. Michie, & R. West (Eds.), *Thinking about behaviour change: An interdisciplinary dialogue*. London: Silverback.
- Kelly, M., & Kelly, R. (2018). Quantifying social influences throughout the life course: Action, structure and 'omics'. In M. Meloni, J. Cromby, P. Fitzgerald, & S. Lloyd (Eds.), *Handbook of Biology and Society* (pp. 587–609). Basingstoke: Palgrave Macmillan.
- Kelly, M. P., Kelly, R. S., & Russo, F. (2014). The integration of social, behavioral, and biological mechanisms in models of pathogenesis. *Perspectives in Biology and Medicine*, 57(3), 308–328.
- Kelly, M. P., & Moore, T. A. (2012). The judgement process in evidence-based medicine and health technology assessment. *Social Theory and Health*, 10(1), 1–19.
- Kelly, M., Morgan, A., Ellis, S., Younger, T., Huntley, J., & Swann, C. (2010). Evidence based public health: A review of the experience of the national institute of health and clinical excellence (NICE) of developing public health guidance in England. *Social Science and Medicine*, 71(6), 1056–1062.

- Kelly, M., & Russo, F. (2018). Causal narratives in public health: The difference between mechanisms of aetiology and mechanisms of prevention in non-communicable diseases. *Sociology of Health and Illness*, 40(1), 82–99.
- Kriznik, N., Kinmonth, A., Ling, T., & Kelly, M. (2018). Moving beyond individual choice in policies to reduce health inequalities: The integration of dynamic with individual explanations. *Journal of Public Health*. Under consideration.
- Marteau, T. M., Hollands, G. J., & Kelly, M. P. (2015). Changing population behaviour and reducing health disparities: Exploring the potential of choice architecture interventions. Technical Report 15-0002, Agency for Healthcare Research and Quality. AHRQ publication.
- Newhouse, M. L., & Thompson, H. (1965). Mesothelioma of pleura and peritoneum following exposure to asbestos in the London area. *Occupational and Environmental Medicine*, 22(4), 261–269.
- NICE (2012). The NICE public health guidance development process. <https://www.nice.org.uk/process/pmg5/chapter/introduction>.
- Ozanne, S. E., & Constância, M. (2007). Mechanisms of disease: The developmental origins of disease and the role of the epigenotype. *Nature Reviews Endocrinology*, 3(7), 539.
- Paisley, S., Aananiadou, S., Levay, P., & Kelly, M. (2018). Public health, policy and the evidence-based medicine paradigm: Challenges and innovations in methods of searching. Under consideration.
- Sytkowski, P. A., D'Agostino, R. B., Belanger, A., & Kannel, W. B. (1996). Sex and time trends in cardiovascular disease incidence and mortality: The Framingham heart study, 1950–1989. *American Journal of Epidemiology*, 143(4), 338–350.
- Tannahill, A. (1985). What is health promotion? *Health Education Journal*, 44(4), 167–168.
- White, D., & Pitts, M. (1998). Educating young people about drugs: A systematic review. *Addiction*, 93(10), 1475–1487.
- WHO (2008). *Closing the gap in a generation: Health equity through action on the social determinants of health: Commission on social determinants of health final report*. World Health Organization.
- Wilkinson, R. G., & Marmot, M. (2003). *Social determinants of health: The solid facts*. World Health Organization.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



## Chapter 10

# Particularisation to an Individual



**Abstract** In Sect. 7.1, we discussed extrapolation from a study population to a target population. In this chapter, we treat particularisation from a study population to one of its members. In both cases, evidence of similarity of mechanisms plays a crucial role.

Inference from an effectiveness claim involving a whole population to effectiveness in one of its members is of central importance in medical diagnosis, prognosis, and treatment. This mode of inference is often called *direct inference* (Kyburg et al. 2001; Wallmann 2017; Wallmann and Williamson 2017).

The case we discuss here is very simple. Evidence of effectiveness in only one population to which the individual belongs is available. The case in which such evidence for several such populations is available is much more complicated and we will not deal with it here. If one has established effectiveness in a population, then one has also established that there is a mechanism operating that connects the putative cause and effect. Now, the population may not be entirely homogeneous with respect to this mechanism: some individuals will exemplify the mechanism while others may not. One way to establish that mechanisms in the population are applicable to a particular individual is by assessing how homogeneous the population is with respect to the mechanism of action. Inference from a homogeneous population to individuals is more likely to succeed, because most individuals will exhibit the mechanism responsible for causation in the population.

However, in most cases there will be subpopulations for which effectiveness does not hold. There may be several reasons for this kind of *exceptionality*. Firstly, in some such subpopulations the mechanism responsible for effectiveness in the whole population simply does not operate. For instance, while drinking considerable amounts of milk is normally safe, subpopulations with lactase deficiency should drink only small amounts of milk. Considering whether crucial features of the mechanism responsible for effectiveness are present in the particular individual can therefore increase certainty about whether the causal claim is applicable to the individual. Secondly, counteracting mechanisms may operate in some subpopulations. For instance, exercising is normally beneficial for preventing stroke by lowering blood cholesterol, but smoking may counteract these beneficial effects by raising blood cholesterol. With this in mind, the following questions can assist the evaluation of evidence of mechanisms for direct inference:

### **Particularisation to an individual**

What is the status of the claim that the mechanism of action in the population is responsible for effectiveness in the individual? Consider the following questions; can both be answered in the affirmative?

**Exemplification.** Are the crucial features of the mechanism of action in the population preserved in the individual?

**Masking.** Are there further mechanisms operating in the individual that counteract the mechanism operating in the population?

When ruling out masking, one needs to pay attention to co-morbidities, social mechanisms, genetic susceptibility and many more. For instance, when assessing whether a certain patient with breast cancer will benefit from a treatment by trastuzumab, one needs to test for HER2. HER2 if overexpressed, increases cell growth over its normal limits. Trastuzumab blocks the effects of overexpression of HER2. If the patient does not overexpress HER2, the drug will not work for her (Bange et al. 2001). Note that if exemplification has been established and masking ruled out, it is possible to particularise a population-level causal claim to an individual without the need for the population to be homogeneous with respect to the mechanism of action. On the other hand, a high degree of homogeneity provides *prima facie* evidence for exemplification and against masking, and thereby supports particularisation.

**Example. Lactose intolerance**

The world population is not very homogeneous with the reaction to milk intake. About 65% of people are lactose intolerant at some point in their lives. However, in different populations there are differing frequencies of lactose intolerant members. Only 5% of Northern Europeans and more than 90% in some populations in East Asia are lactose intolerant, for instance (NIH 2017). This is because in East Asia lactase deficiency is quite common, while it is quite unusual in Northern Europe. Now, establishing that the patient has no lactase deficiency may be sufficient to establish that she may safely drink milk at high doses. However, even if ruling out lactase deficiency is not possible, establishing homogeneity in a relevant subpopulation may provide grounds for provisionally establishing causality in its members. If, for instance, a patient is North European, this may make it quite plausible that she can drink milk safely. If, on the other hand, a patient is East Asian, this may make it quite plausible that she cannot drink milk safely.

**Example. The Shonubi case**

Nigerian drug-mule Shonubi was caught on his eighth trip from Nigeria on the JFK airport carrying heroin in his digestive tract (Colyvan et al. 2001). For sentencing purposes, it was assessed whether the total amount of drugs smuggled on his seven prior trips was greater than a specific amount  $M$ . There was statistical data available for the amount of drugs carried by balloon-swallowing heroin smugglers from Nigeria. Moreover, there is a social mechanism involving these smugglers that helps to explain the amount of drugs they smuggle: the local drug organisation trains the mules in balloon-swallowing for several weeks and threatens people who refuse with violence (Izenman 2000).

It seems best to estimate the amount of drugs smuggled by Shonubi on his seven prior trips by the average amount smuggled by balloon-swallowing heroin smugglers from Nigeria. There is high quality mechanistic evidence for application to Shonubi available. Firstly, the mechanism that connects balloon-swallowing heroin smugglers from Nigeria to the quantity of drugs smuggled does apply to Shonubi. The local organisation did indeed train Shonubi by similar methods to those applied to other drug mules, for instance. Secondly, it seems that, for all we know, there is no counteracting mechanism that makes Shonubi an exceptional drug mule. Note that the trip on which he was caught was already his eighth. Thirdly, although there is some variability with respect to the amount smuggled within balloon-swallowing heroin smugglers from Nigeria, virtually all drug mules smuggled more than  $M$  grams. Hence, the balloon-swallowing heroin smugglers from Nigeria is arguably a sufficiently homogeneous population.

**Table 10.1** Determining the status of the causal claim in the individual given the status of the causal claim in the population and the status of the claim that the mechanism of action in individual and population is similar

		Similarity of mechanism in individual and population				
		Established	Provisionally established	Other cases	Provisionally ruled out	Ruled out
Causation in population	Established	Established	Provisionally established	Arguable	Speculative	Speculative
	Provisionally established	Provisionally established	Arguable	Speculative	Speculative	Speculative
	Arguable	Arguable	Speculative	Speculative	Speculative	Speculative
	Speculative	Speculative	Speculative	Speculative	Speculative	Speculative

To obtain the status of effectiveness for a particular individual, one can combine the status of the effectiveness claim in the population with the status of the mechanistic similarity claim (i.e., the claim that there is exemplification and no masking), as in Table 10.1.

A few remarks shed some light on this table.

First, observe that effectiveness in an individual can almost never be ruled out by the fact that the mechanism responsible for effectiveness in the population is not present in the individual. After all, the individual may exemplify an alternative mechanism of action. I.e., the individual may be a member of a different population, which also exhibits effectiveness but with a different mechanism of action, and this alternative mechanism is present in the individual.

Second, particularisation is a special case of extrapolation. When particularised, a causal claim is extrapolated to the subpopulation of population-members that share all the relevant properties of the individual. This target subpopulation will typically be small, but it remains a subpopulation. Suppose, for instance, we are interested in whether a 30 year old Norwegian farmer will develop an adverse reaction when drinking milk. 95% of individuals in Northern Europe show no such reaction. Here, the target population relevant to particularisation may contain only the farmer in question, while the study population is the class of all Northern Europeans.

Third, there are nevertheless some differences between the evaluation of external validity and the evaluation of particularisation to an individual. Particularisation to the individual is more likely to succeed than is extrapolation from a study population to a target population that is not a subpopulation of the study population. This is because causality established in a population is more informative about individuals in this population than about individuals in different populations. For instance, if the population is very homogeneous, then particularisation to the individual is likely to succeed while extrapolation to other populations may well fail. This fact is reflected in the above tables. Consider the case where no studies are available which involve the particular individual. If mechanistic similarity is provisionally

established and effectiveness is established in the population, the causal claim is *provisionally established* for the individual, according to the particularisation table. In the case of external validity, if mechanistic similarity between the study and target populations is provisionally established and effectiveness is established in the study population, effectiveness in the target population is only *arguable* (see Sect. 7.2). It is worth emphasizing here though that particularisation to an individual is still an extrapolation, and should still be considered fallible.

Note finally that, in contrast to the method of evaluating external validity in Sect. 7.2, in the present chapter we treat the case where there is no evidence for causation obtained by studies directly on the target population.

## References

- Bange, J., Zwick, E., & Ullrich, A. (2001). Molecular targets for breast cancer therapy and prevention. *Nature medicine*, 7(5), 548–552.
- Colyvan, M., Regan, H. M., & Ferson, S. (2001). Is it a crime to belong to a reference class? *Journal of Political Philosophy*, 9(2), 168–181.
- Izenman, A. J. (2000). Assessing the statistical evidence in the Shonubi case. In J. Gastwith (Ed.), *Statistical science in the courtroom* (pp. 415–443). New York: Springer.
- Kyburg, H., & Teng, C. M. (2001). *The Theory of probability*. Cambridge: Cambridge University Press.
- NIH (2017). Lactose intolerance: US national library of medicine, genetics home reference. <https://ghr.nlm.nih.gov/condition/lactose-intolerance>.
- Wallmann, C., & Williamson, J. (2017). Four approaches to the reference class problem. In G. Hofer-Szabo & L. Wroński (Eds.), *Making it formally explicit: Probability, causality, and determinism*. Dordrecht: Springer.
- Wallmann, C. (2017). A Bayesian solution to the conflict of narrowness and precision in direct inference. *Journal for General Philosophy of Science*, 48, 485–500.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

