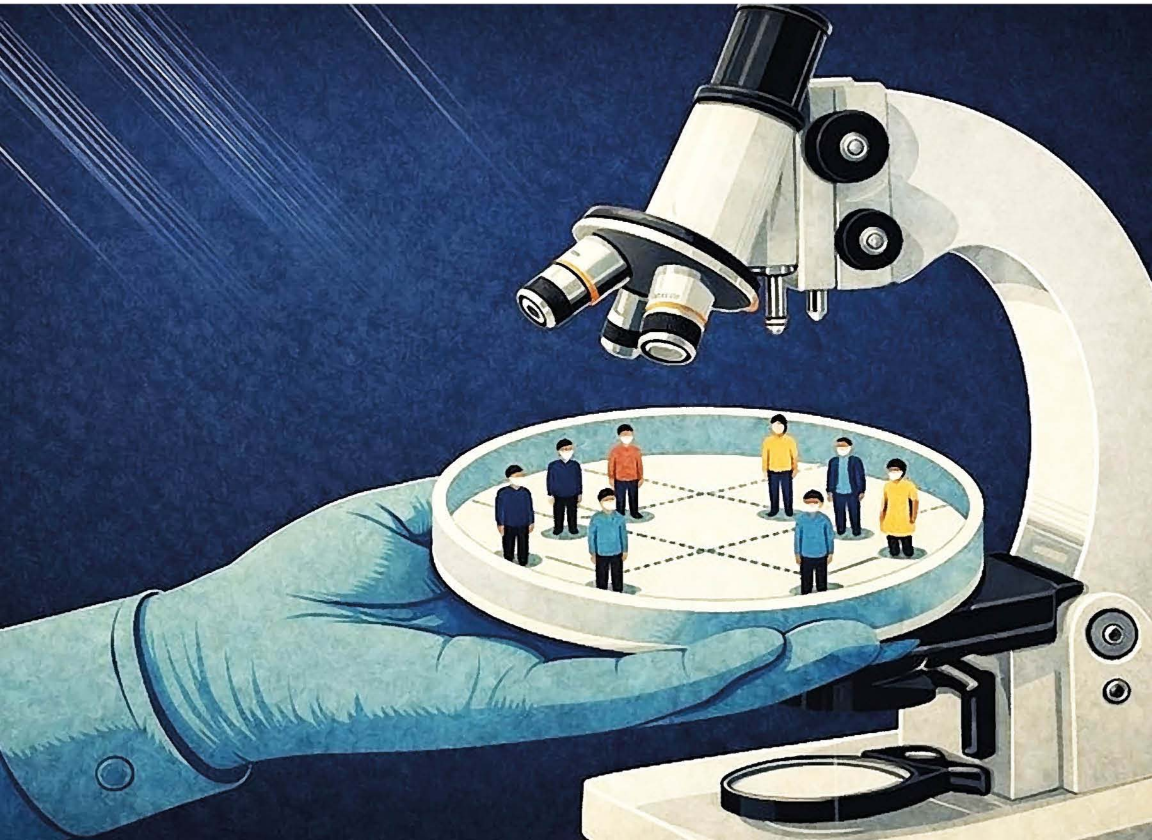
Power, Knowledge, and Covid-19

The Making of a Scientific Orthodoxy

ALEX BROADBENT AND PIETER STREICHER



POWER, KNOWLEDGE, AND COVID-19

Power, Knowledge, and Covid-19: The Making of a Scientific Orthodoxy shows, step by step, how a dominant scientific line on Covid-19 was built and defended – and what it left out.

Through tightly argued case studies, Alex Broadbent and Pieter Streicher reconstruct how early modelling distinctions (notably the suppression/mitigation frame) and threshold-based reasoning made lockdown the default; how debates on masking and vaccination hardened into dogma; and how rival views were sidelined through credentialing, gatekeeping, and the control of forums. The book names and analyses five recurring features of this orthodoxy – methodological rigidity, scientific dogma, suppression of dissent, indirect political authority ("follow the science"), and scientific injustice – and shows how each shaped decisions across diverse settings.

Pairing clear conceptual analysis with accessible evidence reviews, the authors probe where models misled, where uncertainty was overstated or understated, and where costs, context, and equity were neglected – especially in low-resource settings. Rather than relitigating the pandemic, they offer a practical framework for recognizing when science and policy converge too tightly, how to keep plurality alive under pressure, and how to design governance that preserves expertise without closing down legitimate choice. For readers in philosophy, public health, policy, and beyond, this is a concise, non-polemical account of what went wrong, what went right, and how to do better next time.

Alex Broadbent is Professor of Philosophy at Durham University, Visiting Professor at the University of Johannesburg, and Director of the Centre for Philosophy of Epidemiology, Medicine, and Public Health. His previous books include *Philosophy of Epidemiology* (2013), *Philosophy for Graduate Students* (2016), and *Philosophy of Medicine* (2019), and he has also edited *The Routledge Handbook of Philosophy of Public Health* (2023, with Sridhar Venkatapuram) and *The Oxford Handbook of Philosophy of Medicine* (2025).

Pieter Streicher is Research Associate at the Centre for Philosophy of Epidemiology, Medicine, and Public Health at the University of Johannesburg and Durham University.

POWER, KNOWLEDGE, AND COVID-19

The Making of a Scientific Orthodoxy

Alex Broadbent and Pieter Streicher

Routledge
Taylor & Francis Group
NEW YORK AND LONDON

Designed cover image: Maryke Sutherland

First published 2026
by Routledge
605 Third Avenue, New York, NY 10158

and by Routledge
4 Park Square, Milton Park, Abingdon, Oxon, OX14 4RN

Routledge is an imprint of the Taylor & Francis Group, an informa business

ISBN: 978-1-041-22485-3 (hbk)
ISBN: 978-1-041-22482-2 (pbk)
ISBN: 978-1-003-73249-5 (ebk)

DOI: 10.4324/9781003732495

Typeset in Times New Roman
by Apex CoVantage, LLC

We dedicate this book to careful thinking in difficult circumstances.

AB, PS

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

CONTENTS

FIGURES

TABLES

ACKNOWLEDGEMENTS

This book and the work that went into it have benefited from wide and varied exposure. We are grateful to all who helped improve the work, whether or not intentionally, and whether or not they accept our arguments or agree with our conclusions. Inclusion below should not be taken to imply support or agreement, but we are nevertheless grateful for the time, energy, and effort that many people have devoted to communicating with us one way or another. At the same time, we apologize for any omissions.

Many of the ideas in the book were informed by collaborations with others, and we would like to single out Benjamin Smart, Herkulaas Combrink, and Joel Hellewell for the part their thinking and efforts played in developing the framework eventually presented here. We would also like to thank a number of other co-authors, including Abejide Ade-Ibijola, Kalipso Chalkidou, Amanda Glassman, Charis Harley, Ebrahim Momoniat, Likhwa Ncube, Farai Nyabadza, Nazarius Tumwesigye, Richard Sullivan, Damian Walker, and members of the United Nations South Africa Covid-19 Rapid Needs Assessment team.

We are grateful to all those who have read and commented on written material, including Nancy Cartwright, Naja Hulvej Rod, Oliver Johnson, Elisabetta Lalumera, Mat Mercuri, Robert Northcott, Jeremy Simon, Christine Stabell Benn, Briana Toole, Eric Winsberg, and Simon Wood. There were also some who asked to remain anonymous.

We benefited from a number of email exchanges concerning the topics of the book, for which we would like to thank Stefan Baral, Lisa Brouwers, Seth Flaxman, and Harry Moultrie. For further helpful conversations (including on social media), we are grateful to Kevin Bardosh, Alexander Bird, Olaf Dammann, Neil Ferguson, Jonathan Fuller, Sandro Galea, John Gillott, Peter Godfrey-Smith, Glenda Gray, Jesse Hamilton, Ichiro Kawachi, Sean Muller, Neil Pearce, Rik Peels, J.M. Pienaar, Peter Singer, Ragnar Van Der Merwe, Jan Vandenbroucke, Daniel Veldsman, Sridhar Venkatapuram, S. V. Subramanian, Peter Vickers, Eric Winsberg, and many others whose names we apologize for failing to record.

Ideas were presented to, and thus refined by, audiences of many kinds, and we are grateful to those at the Bioethics Meets Philosophy conference at Jagiellonian University; Boston School of Public Health; the British Society for the Philosophy of Science; Cambridge University History and Philosophy of Science Department; Cambridge University Moral Sciences Club; Cambridge Virtual Conversations; New York University School of Law Center for Human Rights and Global Justice; Durham University Centre for Humanities Engaging Science and Society; Durham and Johannesburg Universities Centre for Philosophy of Epidemiology, Medicine and Public Health; Durham University Institute for Medical Humanities (including the Wellcome Discovery Research Platform); Durham University Philosophy Department (especially students in Philosophical Issues in Contemporary Science 2022–24); Jagiellonian University; Korean Society for Preventive Medicine; Korean Society of Philosophy of Science; Liverpool University Philosophy Department; Philosophy of Science Association; Princeton University Center for Human Values; King's College London Philosophy Department Sowerby Project; The New Institute Hamburg; Tufts University School of Public Health; Boston College School of Public Health; University College Groningen; University of Athens; University of Johannesburg Department of Philosophy; University of Macau; University of South Bohemia; and Virginia Commonwealth University.

This research was funded in part by the Wellcome Trust Grant number 226798/Z/22/Z, which supports the Discovery Research Platform for Medical Humanities at Durham University. Much of the book was written at the Platform, which provided considerable support, both institutional and intangible. The Department of Philosophy and Faculty of Arts and Humanities both provided seedcorn funding for aspects of this project, and the Department of Philosophy at the University of Johannesburg and the SOCRATES Centre in Hannover both hosted key research. We are grateful to all these institutions for their part in providing time, resources, and environments congenial to research and writing.

Alex would like to acknowledge the support of family, friends, and colleagues throughout the Pandemic and the development of this book. Piet would like to acknowledge his wife, Ronel, and family for their patience and understanding during a period when engagement with Covid-19 – and subsequently this book – became all-consuming. Their support and tolerance made it possible to sustain the focus and persistence that this work demanded.

Notwithstanding our debts to all the aforementioned, the views expressed in this book, and any remaining errors, omissions, and misjudgements, are ours alone.

INTRODUCTION

This book is a sense-making exercise: an effort not only to understand, but also to describe – to understand through describing, or redescribing. During the Covid-19 Pandemic of 2020–23 (henceforth "the Pandemic"), science had a dramatic impact on policy: larger than ever before, we suspect, certainly if one considers the extremely short timeframes involved. Why? What was the impact? How did it come about? Was it right that some scientists enjoyed a very significant influence for some periods of time? What explains the attacks that scientists sometimes launched upon each other, and does anything justify them? Most people feel something went wrong, but there is great disagreement about what. Some feel scientists overstepped the mark considerably and were instrumental in very damaging policy decisions. Others feel scientists were grievously scapegoated for the mistakes of politicians, and that if they had been heeded sooner, millions of lives would have been saved. As the Pandemic recedes into the past, neither view has emerged as the standard. The danger is that when a similar situation arises, we will be no better placed to manage it than we were in 2020: not because we do not have stockpiles of vaccine, but because we do not know, as a global society, how to properly manage scientific advice in emergencies.

This book is an extended answer to the question, "What went wrong at the science-policy interface during the Pandemic?" Our answer is that there was a *scientific orthodoxy:* a politico-epistemic system in which there is methodological rigidity, scientific dogma, suppression of dissent, illegitimate indirect political authority, and scientific injustice. In this book, we provide definitions of each of these elements that can serve as "tests" to detect them and thus prevent the same thing from happening again.

DOI: 10.4324/9781003732495-1

It must be stressed that this is not an "anti-science" book. On the contrary, we rely throughout on scientific material, and we argue that, sometimes, scientists have the legitimate political authority to direct policy makers – to tell them what to do. This is in complete contrast to the standard view of scientific advice, which is that it should lay out options. In our view, there are times when the scientist is not only permitted but required to go well beyond laying out options and must recommend or direct. Sometimes, the science *should* be followed. But only sometimes: only in specific situations, when certain conditions were met. Yet what are these conditions that determine whether a situation is one in which the science should be followed? During the Pandemic, some were convinced that this was such a situation, as is clear from their making strong, all-things-considered policy recommendations. However, there was no clarity on how to decide whether the situation was one in which science should be followed. This unclarity persists, and dealing with it is one major task for the present book. If the science is to be followed *sometimes,* it is vitally important to be clear about when.

One big challenge in writing about the Pandemic, and about the role of science in society generally, is identifying "science." However, our challenge in this book is slightly different from the philosophical challenge of saying what science is in general. Our challenge is to identify the actor or actors who happened to receive that label during the Pandemic. It is not always easy, and one of the most striking conclusions we reached, in thinking about the Pandemic, is that the people and institutions who came to speak with the voice of science were a very small, select group. When the phrase "Follow the science" became popular in 2020, the science in question was infectious disease modelling, a small part of epidemiology, itself a relatively small scientific discipline. Only one of the World Health Organization's 800 collaborating centres worked on infectious disease modelling. That centre was, therefore, incredibly influential for a time, in a way that was probably not foreseen by anyone, and we suspect it is probably still not widely understood. Of course, the situation evolved, but the dynamic, whereby a relatively small configuration of scientific authorities exerted outsized control over what "science" was thought to "say," persisted throughout the Pandemic, and to some extent remains with us at the time of writing, as we shall argue. The idea of an orthodoxy is meant to capture this situation, where a small group or a single institution comes to stand almost for the whole of science. An ecclesiastical orthodoxy is not merely a system of propositions, but a structure that empowers certain institutions or individuals with the role of epistemic authority. So it was with the Pandemic Orthodoxy, we argue.

This book is grounded in the philosophy of science, but it is not a traditional philosophy book, because its primary goal is not to advance our understanding of philosophy. The goal is not to use the Pandemic to do philosophy, but to use philosophy to understand the Pandemic. Our contribution is thus part of the recent *applied turn* in the philosophy of science.

In making the applied turn, philosophy of science seeks to address philosophical problems that arise in and around science when it is done and used. Traditionally, philosophy of science still drew its problems from real-world concerns, such as distinguishing science from pseudoscience. But it treated them with a degree of abstraction that has come to feel unsatisfying to many contemporary philosophers. The solutions and debates took on a life of their own, and their application to the context from which they arose came to seem distant, or forgotten, like a student essay that rehearses the relevant material but forgets to answer the question. The applied turn is often celebrated as an advance. But, with an eye to avoiding Whiggishness, it may be wiser and more humble to see the applied turn simply as a development, or perhaps an adjustment to the intellectual and sociopolitical climate of the day.

Solving is never on the cards for philosophical problems, and claims to have solved problems should be treated with suspicion. The test for applied philosophy is whether it identifies philosophical problems that were causing trouble in the target context, whether it helps to work with them or round them, and, if possible, whether it improves the prospects for dealing with other or future instances of the same or similar problems. Our topic is the relation between science and policy in the Pandemic, and our problem is that some things went wrong. Unjust policies were implemented (we argue); poor predictions were made (we show); and unseemly disputes abounded (everyone knows). Without a framework, one is apt to end up with a list of "a hundred and one things about Covid-19," useless for any purpose other than recrimination. Our proposal that a scientific orthodoxy emerged during the Pandemic is meant to organize and explain the problems and avoid similar problems emerging in the future.

1
SCIENTIFIC ORTHODOXY

1.1 Introduction

The central claim of this book is that, during the Pandemic, a *scientific orthodoxy* arose. The argument of the book thus has two purposes throughout: to define that concept, and to show that it applies to the Pandemic.

We are prompted to develop a new concept by the *strangeness* of what happened. Why was "follow the science" such a powerful directive? And when things went wrong, as everyone admits they did at least sometimes, was this because the science was not followed, or because the science was wrong? In the years that have elapsed since the Pandemic, and since the writing of this book began, opinions on these questions have mellowed, but not reconciled. There is no social consensus on questions such as whether lockdowns were necessary, whether masks were effective, or how widespread vaccination against Covid-19 ought to have been. This is partly because it is unclear whether such questions should be understood as political or scientific, and thus how to answer them, or who to listen to when seeking an answer.

The slogan "follow the science" provided an answer to the question of who to listen to, but it was a difficult answer for many to fully accept. For one thing, the kind of decisions being made – about whether to implement nationwide lockdowns, for example – involve many considerations that lie in the political domain, such as how to balance the needs and rights of different people. At a deeper level, science seeks to distinguish itself by an attitude of critical inquiry, at odds with the idea of "following." As many scientists see it, modern science began with the rejection of ecclesiastical and scholastic authority and

DOI: 10.4324/9781003732495-2

distinguishes itself by its lack of articles of faith. Setting science up as a new absolute authority goes against the grain for many scientists, even if there are sociologists and philosophers who claim that science is no less authoritarian or faith-based than what went before (Bloor 2007).

On the other hand, a complete rejection of scientific authority is out of the question. No reasonable contemporary state can operate without giving science a special seat at the table in making decisions about almost anything, because science is the major knowledge-producer in contemporary society. To deny science a special place at the decision-making table is ridiculous. Yet in the Pandemic, science seemed to have the driving seat, or to want it – or at least some scientists seemed to, sometimes. Hence the question: when things went wrong, was it because the science was followed, or because it was not? What went wrong at the interface between science and policy during the Pandemic?

Our answer seeks a middle ground, somewhere in between blaming science and exonerating it. We argue that an unusual dynamic arose, which we call *scientific orthodoxy*. This resulted in scientific advice that was unusually forceful and at the same time error-prone, and for essentially the same reasons: that it was methodologically rigid, dogmatic, intolerant of dissent, illegitimately authoritative, and negligent of the interests of many people. This dynamic is not science itself, but an interaction between science and politics. It is thus not a necessary feature of science, nor even of scientific advice. Nor is the dynamic that arose in the Pandemic to be understood as a set of actions taken by any particular scientists or institutions. It is true that some actors did make strenuous efforts to influence policy, and we criticize some of these. However, there are always scientific advocates in the policy arena (Pielke 2007, 3). What calls for explanation is the fact that some scientific actors had so much *influence* at certain points during the Pandemic, often to the exclusion of other scientific actors. This is why we postulate a systemic phenomenon with political properties, identical neither with science in general nor the actions of any particular scientific actors: we postulate a scientific orthodoxy.

1.2 Why Something New?

Philosophers, sociologists, historians, anthropologists, and even scientists themselves have thought and written a great deal about the way science interacts with the society that embeds it. Does the Pandemic call for a theoretical development, such as we propose? Does it exhibit anything new, not already captured in those rich literatures?

We believe so. There are several reasons, and we elaborate on some of them later in the book, but the main reason is simply that something special happened – something that was *not* science as usual. Compare, for instance, Thomas Kuhn's fabulously insightful analysis of science (Kuhn 1962, 1977). Our approach owes a great deal to Kuhn. Yet Kuhn's theory cannot explain the Pandemic. Kuhn set out to say what makes science special, in common with others of his day. He emphasized a distinction between mature and immature science, and confined his analysis to the most mature: physics and chemistry. He sought to describe, not to prescribe: his was a judgement-free account. He exposed the role of political and social dynamics in science, and especially how they led to theory change – a process he rendered so dramatic that he said scientists working in different paradigms lived in different worlds.

The Pandemic, on the other hand, was strongly influenced by epidemiology, a highly immature science by Kuhnian standards, with multiple theoretical frameworks vying for dominance (as we will see in several chapters of this book). These scientists may have had very different approaches to certain questions, but it is doubtful that they were working in different paradigms, and even more doubtful that a paradigm shift occurred or was in any other way implicated during the Pandemic. Moreover, there was competition between scientific disciplines in the public sphere, something that Kuhn's account does not capture. Kuhn's exposé of the role of social dynamics in science does not tell us how science plays a role in policy, nor whether it should. Kuhn's account is descriptive, not normative; he is interested in accurately saying what scientists do, not what they should do. The whole point of Kuhn's account is to explain what makes science distinctive. He was not in the slightest concerned with saying how scientists ought to advise policy-makers, let alone in an emergency situation such as a Pandemic. Kuhn's influence is obvious throughout this book, but no serious Kuhn scholar would suggest that the events of the Pandemic can be packaged as a mere evolution of some Kuhnian process.

This is because what happened in the Pandemic was *distinctive*. It was a dynamic that arose at the interface between science and policy, but one that is not always there. Science is not always orthodox, in the sense we describe in this book. The episodic nature of what happened is the central reason that existing accounts of science, policy, society, and the interactions between them fall short.

This answer will not satisfy someone whose primary interest is in the lessons of the Pandemic for philosophy, history, sociology, science and technology studies, or any other academic field. Our stance, as we have already explained, is that we are interested in using philosophy and other tools to understand the Pandemic – not using the Pandemic to advance philosophy or other academic disciplines. We develop a new tool because we need it, not because we have an urge to contribute to this or that scholarly literature. One who starts with the literature will want quite a different kind of book – one that starts with the literature. We start with the Pandemic itself. This is not a work that aims to intervene in the philosophical literature by comparing and contrasting theories, evaluating counterexamples, or expanding on conceptual distinctions. Instead, it is a philosophical analysis in a realist and empirically grounded spirit, using conceptual tools to understand a real-world phenomenon that existing literature fails to adequately capture. It is "field philosophy" in the spirit of Nancy Cartwright, Ian Hacking, and many others (Brister and Frodeman 2020). It is motivated by the reality of what happened, not by academic literature either beforehand or in response. We appeal to that literature where it is useful as a tool, but not as an anchor.

1.3 The Elements of Scientific Orthodoxy

We say that a *scientific orthodoxy* arises when there is

(a) Methodological rigidity,
(b) Scientific dogma,
(c) Suppression of dissent,
(d) Illegitimate indirect political authority, and
(e) Scientific injustice.

We now define and explain each of these elements further.

1.3.1 Methodological Rigidity

We define methodological rigidity as follows.

Methodological rigidity is

(a) Reliance on one methodological approach without due consideration of its
(b) Applicability,
(c) Vulnerabilities, or
(d) Track record, or of
(e) Other relevant approaches.

Scientists often face a choice between different methods. Studies do not design themselves, and the design process is full of decisions. Data analysis is even more methodologically flexible, since answering a question in a different way does not require one to conduct a whole new study, merely to make different choices about how to proceed with the analysis. These choices can result in completely different answers to exactly the same questions about exactly the same data (Silberzahn et al. 2018). Policy-makers also face choices between the outputs of different methods, because sometimes one type of research points in one direction and while another type points in another. The most famous instance of this is the tension that often exists between randomized trials and observational research, and this arose in the context of the Pandemic in the context of masks, discussed in Chapter 6. Science is not always univocal, and scientists sometimes disagree; and sometimes, their disagreement concerns method.

Differences of scientific opinion, including in matters of method, can be reasonable. Methodological rigidity is a case where there is an *unreasonable* fixation on one method or one methodological approach. In the Pandemic, this took the form of a fixation on one approach to modelling the transmission and consequent mortality of Covid-19. Broadly speaking, one can distinguish two approaches to modelling the course of an infectious disease (Kucharski 2020). A *mechanistic* approach seeks to capture the underlying dynamics of transmission, and then work out what this will mean for the distribution of disease in the future, or in several alternative futures, depending on what policy decisions are taken. The most sophisticated mechanistic approaches use powerful computers to simulate entire populations. *Descriptive* approaches, on the other hand, seek to characterize the pattern underlying the spread of the disease and then extend this into the future. For example, a modeller might predict what is going to happen in one place by working out where that place is located on the epidemic curve as it played out in another place.

In Chapter 2, we show how some scientific institutions focused exclusively on the mechanistic approach. These institutions were the best resourced and most powerful in the world, explicitly charged by the WHO with developing new methods and teaching them to other scientists. These institutions took action in the public sphere to discredit scientists who used other approaches. The result, we show in Chapter 3, was an inaccurate prediction by some of the best-resourced and most expert epidemiological modellers in the world, over a period of two years. During this time, the majority of modellers quietly switched over to using a mix of methods, with descriptive approaches favoured for short-term forecasting, for good reasons. Those who were methodologically rigid, however, did not assess the reliability, applicability, vulnerabilities, or track record of their models, nor the relevance of other approaches. This, we show, led to error, which eventually became so clear that the groups in question ceased to operate.

1.3.2 Scientific Dogma

We define scientific dogma as follows.

A *scientific dogma* is

(a) A proposition,
(b) Asserted by a scientific authority, and
(c) Inappropriately maintained in the face of recalcitrant evidence.

Dictionary definitions define dogma along the lines of laying down a principle as an incontrovertible truth. Our definition is adjusted to the scientific context by replacing incontrovertible truth with an inappropriate attitude to evidence. This is because science distinguishes itself by its attitude to evidence. Dictionary definitions are not always useful starting points for analytic purposes, but in this case, using one preserves the connection with common usage and thus supports the analogical aspect of our account.

We intend the term "assert" in a colloquial way to indicate more than merely stating a proposition. Philosophers tend to use "assert" as a neutral way of saying that a proposition was said, written, signed, or conveyed in some other way. Colloquially, however, assertion implies something about the attitude and actions of the utterer. One can assert not just a proposition, but also oneself, and one can be assertive. All of these imply, first, an attitude on the part of the utterer, and second, some action taken with the intention of securing acceptance of the proposition. We intend these connotations in our definition of dogma. A scientific authority that asserts a dogma adopts an assertive stance and takes some action to encourage acceptance.

The medieval church used the term "dogma" approvingly. The term has come to be pejorative, of course, and contemporary science would never openly commit to dogma. Moreover, unlike in the medieval church, there is no single organization capable of issuing dogma, and this complicates the situation when a scientific authority wishes (albeit unwittingly) to assert one. This means that we must infer the proposition asserted from a range of statements that a scientific authority makes, and from the claims it takes the trouble to defend.

We may also infer the content of the dogma from the manner in which that defence is conducted. In particular, we focus on resistance to recalcitrant evidence. Some claims were maintained during the Pandemic in the face of evidence that was recalcitrant, either because it contradicted the claim, failed to provide strong support, or provided support only for a weaker version of the claim.

We realize that these are serious allegations to make of any scientific authority – organization, institution, or individual scientist. They are also surprising: they call for explanation. We identify three subjects of dogma: lockdowns, masks, and vaccines. In Chapter 4, we explain how science was involved in the construction of the lockdown strategy, which was where scientific dogma in the Pandemic originated. The source of the Lockdown Dogma, as we call it, was methodological rigidity. The methods favoured by the most powerful and best-resourced epidemiological modelling teams predicted that there was a very sharp divergence between outcomes, depending on whether one succeeded in getting the reproductive number, R_t, below 1. This led them to assert that there were fundamentally two strategic approaches, and that only one of them was acceptable. At that time, the world was confronted with the examples of China, which was thought at the time to have controlled Covid-19, and Italy, which manifestly failed to respond in time and so locked down, showing that lockdowns in liberal democracies were possible. The result was that the fundamental distinction came to be seen in practice as a choice between locking down and not doing so, with non-lockdown options treated as one because they would all lead to unacceptable outcomes.

Once this framework was in place, it became important for the scientific authorities who had constructed the "fundamental" strategic framework to sustain it, since otherwise their authority would be severely tarnished. This explains why their efforts focused exclusively on evaluating the effects of lockdowns, and why they entirely neglected retrospective evaluation of their own predictions. The evaluations of lockdown found that they were both necessary and effective for avoiding very significant mortality. However, as we show in Chapter 5, these assessments were flawed. One study was reported as showing that lockdown had saved 470,000 lives in the UK and over 3 million lives in the EU. However, there was a simple error in the data: the study used the date on which death was reported, rather than the date of death itself. When this is corrected, the model predicts that lockdowns in England occurred after infections peaked, meaning that it cannot have caused them to peak, and thus cannot have been responsible for bringing R_t below 1. In that case, the lives it saved would have been orders of magnitude less numerous. This paper was never corrected publicly, but the authors became aware of the error. A later paper used corrected data, but a different method, and still found that lockdown was effective in bringing about the peak of infections, albeit with a much smaller effect.

The point here is not so much whether lockdowns had an effect, but how the proposition that they were necessary and effective was evaluated. That proposition was maintained in the face of recalcitrant evidence: evidence showing, powerfully, that the effectiveness of lockdown and the question of whether it was necessary were much more complicated, and could not be answered as simply as the Lockdown Dogma asserted. It is the maintenance of the proposition in the face of recalcitrant evidence, rather than the strict literal falsity of the proposition, that makes it a dogma.

Masks and vaccines were also subjects of dogma. Strong, simple assertions were made in the face of evidence that was ambivalent, contrary, or more complicated. Methodological considerations are again implicated in the case of masks. Roughly, epidemiological studies may be divided into *observational studies* and *randomized trials* (sometimes also called *experimental studies, randomized controlled trials,* or simply *trials*). The weight of observational evidence points towards an effect of mask use in the general population. The weight of evidence from randomized trials does not. Given that masks are effective in lab conditions, where does this leave us? One stance is that randomized trials are not necessary to prove effectiveness: they are powerful, but notoriously difficult to apply to complicated real-world situations. Another stance is that observational studies are notorious for failing to eliminate bias, and that, in this case, all the observational evidence is scored at "critical risk of bias" on widely used tools for evaluating such evidence. There is no straightforward solution to this problem. But some scientific authorities issued strong messages that masks were necessary and effective measures against an epidemic wave of Covid-19. We argue in Chapter 6 that this stance was dogmatic.

Vaccines are a fraught political topic, so it is no surprise that they were a point of tension in the Pandemic. We are absolutely clear that vaccination programmes against Covid-19 saved millions of lives. Nevertheless, in Chapter 7, we maintain that vaccines were the object of dogma. Many statements about vaccine effectiveness were subject to the healthy vaccinee bias, a well-known difficulty with assessing vaccine effectiveness in observational studies. This bias operated across the spectrum, spanning public-facing infographics to publications in the *New England Journal of Medicine* (a top-tier medical journal). The use of boosters and the vaccination of children was also pushed in some places to an extent that was not merited either by evidence or immunological theory. Most significantly, some governments and more private entities (employers, in particular) made vaccination compulsory, despite the fact that there is no public interest justification for doing so. This is because Covid-19 vaccines do not offer significant protection against infection (either being infected or infecting others). The primary benefit of vaccination against Covid-19 is that it protects the recipient against serious disease, but not against catching the disease in the first place. Therefore, being unvaccinated does not pose a risk to others, undercutting the usual justification for a mandate.

In each of these cases, we press for greater nuance, more sophistication, and better response to evidence. The existence of scientific dogma in relation to lockdowns especially, but also masks and vaccines, did avoidable harm, some in the form of avoidable mortality and some in other spheres besides health. Dogmatic science does not lend itself to effective public policy.

1.3.3 Suppression of Dissent

There is *suppression of dissent* by a scientific authority when it takes active steps to discredit dissenting views and their proponents.

In Chapter 8, we compare the treatment of some scientists who held non-standard views during the Pandemic to the treatment of medieval heretics. We focus, in particular, on the Great Barrington Declaration (GBD) and on the fourteenth-century English heretic John Wycliffe. We do not defend the GBD, either its content or the manner in which it was put forward, any more than we agree with John Wycliffe. However, there are striking parallels between what Wycliffe did and what the GBD authors did, and between the responses of the ecclesiastical and scientific orthodoxies respectively. Exploring these is fascinating in its own right, but it also illuminates the mechanisms by which an epistemic authority deals with opposition. These mechanisms are antithetical to good science and to good policy.

1.3.4 Illegitimate Indirect Political Authority

Legitimate *indirect political authority* (IPA) exists when:

(a) A person, institution, or other actor has exclusive epistemic access to a domain, and knows a certain fact in this domain [Exclusive Epistemic Access];
(b) There is a background of widely accepted knowledge and values to which no reasonable subject could object [Shared Background]; and
(c) Combining this fact with this background greatly restricts the range of acceptable policy options, potentially to a single one [Keystone Fact].

("Fact" should be read as including the conjunction of multiple facts.)

Illegitimate IPA arises when a person or institution is granted the rights and powers appropriate to legitimate IPA, but at least one of (a)–(c) is not satisfied.

Many would find it counter-intuitive that scientists should ever be in a position to tell politicians what to do, assuming the latter have been elected or duly appointed through a sound process. Probably the most influential view on how science ought to behave in advisory contexts is that it should play the role of an "honest broker" between policy options (Pielke 2007). This is not an entirely neutral role: there are too many decisions to make in selecting which options to inform the decision-maker about (Pamuk 2022), in parameterizing the models (Winsberg and Harvard 2024), and even in selecting the relevant evidence (Fuller 2020). A useful advisor must take into account what the policy-makers are trying to achieve. At the same time, a scientist has no more right than any citizen to make decisions about policies. The advisor should not, therefore, make recommendations based on their own views about which outcomes are more desirable but should, on this view, merely provide information about the consequences of policy decisions. The science advisor needs to consider goals and values in order to be useful, but there remains a line between advice and advocacy, which the advisor ought not to cross – on the most common view.

During the Pandemic, some infectious disease modellers clearly crossed this line. Consider:

> We therefore conclude that epidemic suppression is the only viable strategy at the current time. The social and economic effects of the measures which are needed to achieve this policy goal will be profound. Many countries have adopted such measures already, but even those countries at an earlier stage of their epidemic (such as the UK) will need to do so imminently.
>
> *(Ferguson et al. 2020, 16)*

This report is the subject of much discussion in this book and elsewhere, for good reasons. One of those reasons is that it pushes for a particular policy. This is not just an advisory input for policy-makers to weigh against other considerations. This is clear from the fact that other considerations are mentioned in the passage just quoted. The conclusion amounts to an all-things-considered recommendation.

It would be easy to dismiss this as overreach. However, we do not want to take that dismissive line. As Jonathan Birch has emphasized, the Pandemic was an emergency, and even if it is wrong for scientists to give normatively heavy advice in normal times, these were not normal times (Birch 2021). In Chapter 9, we argue that, sometimes, scientists may be justified in making recommendations that go beyond the usual remit of scientific advice. We compare the situation of epidemiological modellers in the Pandemic to that of Matt Hooper, the marine biologist in the film *Jaws.* Hooper knows there is a dangerous shark in the vicinity and urges the Mayor and Police Chief to take action. He does not lay out options and point out consequences, then leave the Mayor to make the final call. He makes it very clear that the Mayor needs to close the beach. And we, as an audience, are invited to feel that Hooper does nothing wrong, notwithstanding his complete lack of any political standing, in contrast to the Mayor's legitimate right to make decisions. From the political theoretical perspective, Hooper is just a concerned citizen lobbying the Mayor, but as an audience, we feel that he has more right to command the Mayor's attention and to direct the Mayor's decision than a normal citizen. The right stems from what he knows.

Our notion of legitimate IPA is meant to capture the legitimacy that we, at any rate, intuitively feel Hooper possessed. (We also consider, and reject, the possibility that this premise is false, and that no scientist should ever do what Hooper did.) The framework is also meant to articulate the way epidemiological modellers might have felt when they pushed for certain strategies. The phrase "only viable strategy" in the passage recently quoted is key. The scientists did not feel that they were expressing a preference between policy options, but that there really was only one acceptable policy option. Economic costs are mentioned, not as countervailing considerations, but as inevitabilities to be prepared for. This is reflected in our definition of IPA.

We go on to question whether modellers satisfied these conditions. Did modellers have exclusive epistemic access to the domain? Did they attend to contextual factors? Was the fact-value background unobjectionable? Were the facts in question keystone facts restricting policy to only one viable option?

We also discuss the prospects for strengthening governance structures within science and at the science–policy interface, particularly for integrating our test for IPA into practical assessments of scientific advice in emergencies. Current frameworks offered by philosophers and political theorists are not designed with emergencies in mind and tend towards the impractical. We indicate how ours might be used in practice.

1.3.5 Scientific Injustice

Scientific injustice occurs when

(a) Science guides policy, which
(b) Distributes benefits or burdens unequally, where this is
(c) Reasonably foreseeable and
(d) Avoidable

Thus formulated, scientific injustice is a form of social injustice, where social injustice is understood in terms of equality of outcomes, as it usually is. Scientific injustice might be thought of as social injustice caused by science's influence on policy, at least by someone who accepts this concept of social injustice.

The tort of negligence also influences our formulation. Negligence does not require intent to cause harm, or even awareness that one might cause harm. "Reasonably foreseeable" is an objective test in law: it means what would be foreseeable to a reasonable person in the position of the defendant. Whether scientists know or intend that their influence on policy will have unjust results is irrelevant to scientific injustice, in our sense.

In Chapter 10, we show that there was significant scientific injustice during the Pandemic: science-guided policy in a way that distributed benefits and burdens unequally, and this was both reasonably foreseeable and avoidable.

Specifically, we show that lockdowns placed a much greater burden on the global poor and benefited them far less both in absolute and relative terms than the global rich. The median age of a Covid-19 death in the UK in January 2021 was 83, and the median age of the population of sub-Saharan Africa is 19.7. This shows not only that the vulnerability of the population to Covid-19 is likely to be lower, but also that many other things pose major threats to life in Africa; thus, it is highly implausible that the costs incurred by responses to Covid-19 were optimal public health expenditures. The circumstances of the poor are often such as to make social distancing very difficult, and lockdown impossible. One in eight people globally lives in a slum. Over half the global population dies outside a hospital, rising to 80% in low-income countries, so protecting (or even preparing) the healthcare system becomes a less obvious rationale for the short-term pain of a lockdown. And, most pressingly of all for ordinary people, those living even in relatively less extreme poverty often do not possess assets, food stocks in the home, cash, or savings, and work in the informal sector, where there is no pay without work, and no benefits.

The notion of scientific injustice is meant to capture the sense that something was wrong about this, and to give us a framework within which to show that science indeed played a causal role in bringing about socially unjust outcomes. But how on earth did that happen, given that science in general – and certainly many scientists and institutions in epidemiology, public health, and other relevant fields – are apparently committed to quite the opposite: to social justice of some form, or at least to a fairer world? We do not know why it happened, but we have an account of how, and that is provided by the other four elements of scientific orthodoxy and explained in the rest of the book.

1.4 The Origin Question

Of the many things we do not discuss, the question of Covid-19's origin stands out. The epicentre of the first outbreak in Wuhan was in an area containing both a wet market and a biological research facility studying coronaviruses. This raised two possibilities: that the novel virus evolved naturally in animals and jumped to humans in the crowded market, or that it was developed during the course of research in the lab and then (presumably) leaked out. There was initially a strong push by scientists, mainstream media, and many politicians to shut down the "lab leak theory" – the claim that the virus originated in a lab. Subsequently, it became clear that this push was in certain respects *too* strong. In particular, it was too strong to present the matter as a purely scientific one. There was no decisive epidemiological evidence linking the outbreak to either the lab or the market, and evidence from the nature of the virus itself was the subject of debate by credible virologists. Setting aside the substance of the question, there was prima facie overreach by people claiming to speak in the name of science. There were also efforts to stifle dissent. These things might make it seem that the question of Covid-19's origin ought to be part of our discussion in this book.

Nevertheless, we do not discuss the question of Covid-19's origins in this book. This is because there are important differences between the origin question and the topics we discuss, even if there are also similarities. We discuss situations where science plays a directive role in making policy. The question of Covid-19's origin is a political and social dispute about science, but not because science is telling people what to do. If the lab leak theory is true, then scientific research created a danger that it failed to control, and caused major loss of life. This taps into an old fear about science (think Frankenstein) and knowledge more generally (think Adam and Eve). It is interesting as well as practically important. However, contrast that with what would be the case if it were true that lockdowns "didn't work" in some important way: for instance, that the first English lockdown was unnecessary to protect the NHS. If that is true (which we argue it is), then scientific *advisors*, not merely researchers, caused unnecessary suffering and death as a result of an error. This is a completely different situation, morally, politically, and scientifically. Morally, the question is affected by the presumed intention, which was to save lives (and thus laudable), as opposed to gaining knowledge or even developing biological warfare capability (morally more complex). Politically, the question is whether scientists were too forward, or alternatively not forward enough, in pushing for one or another policy, while virologists in the lab in Wuhan were not seeking to alter policy at all in any direct way. Scientifically, the fields are completely different: public health is essentially public and scientifically heterogeneous, while the kind of research done in the lab in Wuhan is often secretive and scientifically highly specialized.

So, while a larger discussion of the politics of science and Covid-19 would surely include the origins question, our specific interest in the role of science in shaping policy during the Pandemic means that we leave it aside.

1.5 Applicability and Flexibility

We intend our account to be widely applicable in both practical and theoretical contexts. Practically, we intend it as a means to detect scientific orthodoxy and a guide to preventing or opposing it. Theoretically, we intend it as a means to understand the relation between science and policy in certain situations. This means we do not see the conditions above as individually necessary or jointly sufficient in the strict sense that philosophers usually have in mind. We are open to the possibility that scientific orthodoxy may take other forms.

Instead of necessary and sufficient conditions, we intend the elements and sub-elements of scientific orthodoxy as analogous to legal tests. Law students learn the rule, the exception to the rule, and the exception to the exception. Likewise, we are open to the possibility that in some situations a scientific orthodoxy may arise without all of our conditions being satisfied. Thus, finding that one of our elements or sub-elements is not necessary does not mean our project was in vain, any more than a legal test (e.g. the "but for" test for factual causation) becomes pointless when exceptions are identified (e.g. harm by multiple tortfeasors).

The theory as a whole, we think of as a kind of "open source" project, akin to open source software. An open source programme permits improvement, adaptation, and extension. It is not a failing of an open source project, but a success that it can be adapted to new uses. So, if our conditions fail to be jointly sufficient for scientific orthodoxy, we are nonetheless satisfied if the approach provides the basis for something else that is useful. In this, again, the guiding analogy of ecclesiastical orthodoxy may prove helpful. Figure 1.1 illustrates the theory as a whole, broken down into each of its elements and sub-elements, and is intended to support the application of scientific orthodoxy to other situations besides the Pandemic.

1.6 Conclusion

A scientific orthodoxy is a dynamic that sometimes arises at the interface between science and policy. It is characterized by methodological rigidity, scientific dogma, suppression of dissent, illegitimate IPA, and scientific injustice. The balance of the book is devoted to illustrating the concept with reference

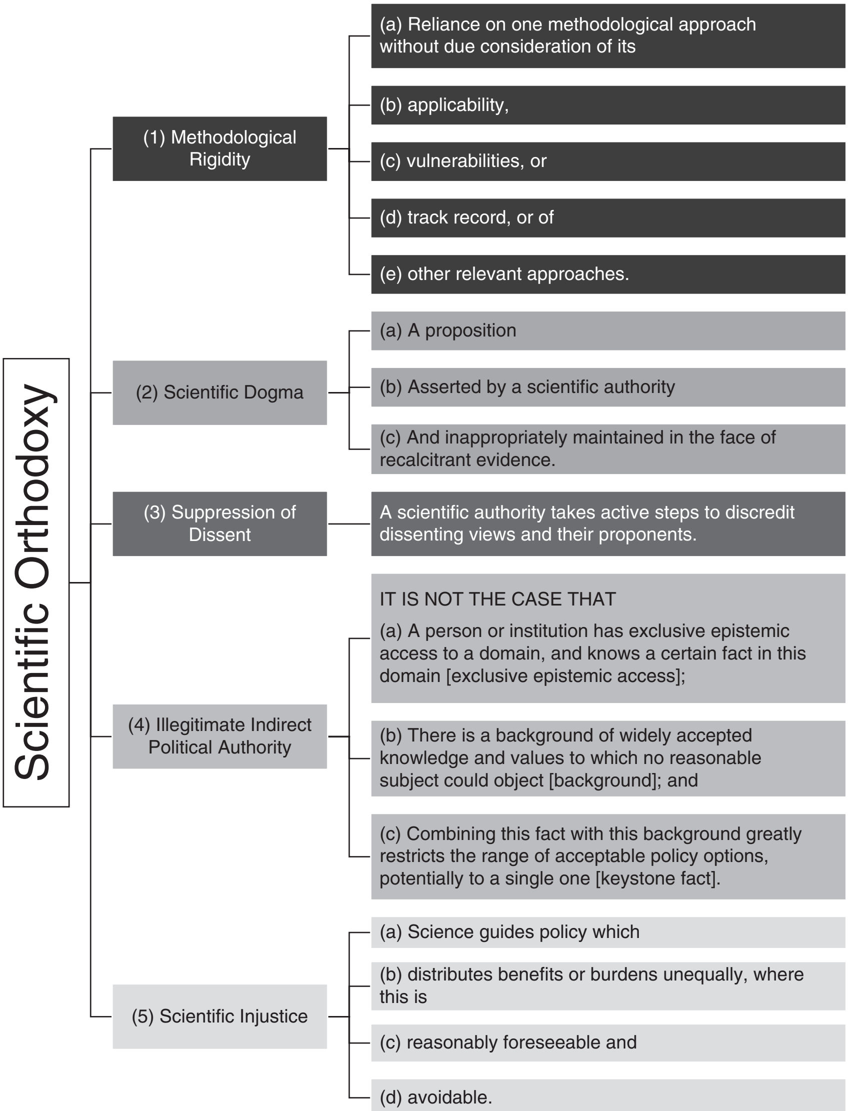


FIGURE 1.1 Scientific Orthodoxy

to the Pandemic, and simultaneously to understanding the process by which a scientific orthodoxy came about. The case informs the theory and the theory illuminates the case, or so we hope.

This project is a contribution to the larger process of preparing for future pandemics. An infectious disease pandemic greatly increases the probability of a scientific orthodoxy emerging. Quick action is required under uncertainty, and fear is present. Scientists are more likely to become more dogmatic, presenting a façade of consensus. For people who respect science, the dynamics that emerged during the Covid-19 Pandemic were disconcerting. In normal science, the goal is to increase knowledge, at least in the eyes of a non-scientist but scientifically favourable public. Different approaches and a contention between ideas are encouraged in this common positive image. But this science-favourable image was often undermined during the Pandemic. It came as a shock to many that science can also be autocratic and exclusionary. Not everyone reached this conclusion, of course; some already saw science this way, while others either were unaware, excused what they did not like, or regarded it as a necessary evil. But there remain some people whose faith in the integrity of science was shaken by what they saw of it. To that extent, the experience of science during the Pandemic alienated some proportion of the population that would otherwise have remained more favourable. By isolating a certain dynamic at the interface of science and politics, and showing that what happened there was unusual, we hope to exonerate the larger scientific project. To put it glibly, the book defends science done right from science done wrong.

Understanding scientific orthodoxy is important for practical and philosophical purposes. The practical purpose is to prepare for future circumstances in which science shapes policy. Perhaps, too, other orthodoxies operating in the present may be discovered. These practical consequences are what we care about most. But the philosophical project of elaborating the nature of an orthodoxy is necessary for accomplishing the practical project. This is an instance where philosophy matters, and, in keeping with many contemporary philosophers, we suspect such instances are more common than either our forebears or people outside the discipline imagine.

2

THE POLITICS OF METHOD IN EPIDEMIOLOGY

2.1 Introduction

It is a familiar misconception that “the scientific method” is a single, unitary thing. Scientists use many different approaches to solve the problems they face, partly because they face different problems, and partly because different methods seem to them to be more appropriate. Differences in opinion about method are difficult to resolve, since there is no direct way to test a method: one must either argue for it from a conceptual standpoint, or identify a track record that shows the method’s general reliability.

This chapter focuses on a particular episode during the Pandemic, which showed how different approaches to epidemiological modelling were competing, not just scientifically, but politically. The two approaches are known, respectively, as descriptive and mechanistic modelling. The WHO Collaborating Centre for Infectious Disease Modelling at Imperial College London (ICL) developed an extremely complex mechanistic model. On the other side of the Atlantic, the Institute for Health Metrics Evaluation (IHME) produced a descriptive model. It differed from ICL’s, not only in its predictions, but in ways that more seriously threatened ICL’s methodological approach. This prompted an attack: a critical response in a scientific journal, coordinated with an apparent public relations campaign to discredit the IHME’s model, and any who would associate with it. The suggestion was that the IHME’s model was irresponsible, and was encouraging governments – especially the American one – to take an unduly relaxed approach to responding to Covid-19. But we show that the scientific grounds for this stance were flimsy; in fact, the IHME’s approach was

DOI: 10.4324/9781003732495-3

subsequently vindicated. The attack took place because of adherence to a certain methodological approach.

This may sound implausible, since it is doubtless not a good account of the motives for the attack. The motive was stated: concern that the IHME's model would not prompt drastic action. The reason for this, however, was not because the IHME was predicting cheerful outcomes, as we explain in the chapter. On the contrary, it was predicting many thousands of deaths, and was initially seen – in the *right-wing* press – as showing a need for restrictions. But it did not show a sharp divergence between different response strategies. This, as we shall explain, is an artefact of the mechanistic models favoured by ICL, which, to look ahead to Chapter 4, we will argue gave rise to a threshold logic, creating the strategic context for lockdown.

2.2 Two Approaches to Infectious Disease Modelling

There are different approaches to infectious disease modelling, and a dispute arose as to which was more appropriate. Approaches to infectious disease modelling can be divided roughly into the following two broad categories (Kucharski 2020).

Descriptive or statistical models attempt to capture the statistical relationships between epidemiological parameters and observed data, without trying to model the physical, biological, or other processes underlying them. The predictive power of the model takes precedence over the extent to which the structure of the model corresponds to the generative mechanism of the data.

Mechanistic models attempt to represent the physical, biological, or other processes of disease transmission. Predictions are derived from modelling these dynamics as they develop in the future.

There are different techniques within these broad approaches, and a particular modelling exercise will often incorporate more than one approach. Nonetheless, this distinction between general approaches to modelling the course of an infectious disease represents an important difference.

Mechanistic models allow the modeller to model alternative scenarios. If the model captures the underlying dynamics adequately for the purpose at hand, then, by adjusting parameters, the model can be made to represent alternative scenarios. Mechanistic models are thus capable of modelling hypothetical, counterfactual, or "what-if" scenarios. The most sophisticated current mechanistic models are *simulation* models, which ideally create a digital twin population of an actual population, and make predictions by seeing how the disease behaves in that twin population. This holds up the prospect of accurate evaluation of different policy options, both in advance, by evaluating the consequences of various options, and in retrospect, by modelling what would have happened if another choice had been made. On the other hand, they rely on simplifying assumptions, which tend to render them less useful for short-term forecasts, as we shall see shortly.

Descriptive models typically contain much less causal information. They express statistical relationships between variables of interest, but do not represent the mechanisms behind these associations (or at least not to a large extent). A simple descriptive model might look at infections for the past week and extend the curve three days into the future. Another might look at the shape of a wave in one location and then seek to determine where another location is located on a wave of the same shape, and thus predict how infections will behave in the future. The ability of a descriptive model to *directly* inform us about counterfactual scenarios is much more limited, because there is much less reason to think that changing the value of a parameter will have a realistic result. The relationship between the variables is statistical but not necessarily causal. Changing the value of some parameter may give a mathematical result, but beyond simple adjustments, there is little reason to think that this bears any relation to the causal effect of altering the real-world feature corresponding to that parameter.

2.3 The Mechanistic Turn

During the Pandemic, mechanistic models were strongly emphasized by the Pandemic Orthodoxy, and descriptive models were actively denigrated. However, this episode, which we will turn to in the next section, did not occur in historical isolation. A move away from statistical models and towards mechanistic models is evident already prior to the Pandemic.

The underlying reason appears to be the ability of mechanistic models to deal with counterfactuals, directly and internally to a single model. The question of what will happen when we intervene then seems to become tractable scientifically.

For example, in a book about the wide applicability of infectious disease epidemiology to many phenomena, Kucharski presents a shift away from descriptive modelling towards mechanistic modelling (Kucharski 2020). He describes the famous historical episode through which SIR models were developed, in which a colonial British medical doctor, Ronald Ross, worked out that one could eliminate malaria in humans by reducing mosquito populations without completely eliminating them. This is a predictive exercise regarding the result of an intervention, and it is not something that a descriptive approach could do (at least not readily, and not from the information available to Ross). Kucharski sees the episode as illustrating simultaneously the power of mechanistic models to predict the outcomes of interventions and their power to yield understanding – their explanatory power. He is in the company of some philosophers here, who have argued that to explain is to provide information about the outcome of possible interventions.

Be that as it may, the inference from the importance of this use of models to their superiority over descriptive models is obviously a non-sequitur. Understanding is not our only goal. Sometimes we want to know what will happen: we want a forecast.[1] Sometimes we want both understanding and a forecast, but sometimes we really just want the forecast. If I am considering a picnic on the beach, I want to know how likely it is to rain, but unless I am in a curious mood, I do not care why.

Interestingly, Kucharski interprets policy-makers as seeking *understanding* from models. This, as we have said, is something that mechanistic models are usually better able to provide than descriptive models. He writes:

> According to Chris Whitty, now the Chief Medical Officer for England, the best mathematical models are not necessarily the ones that try to make an accurate forecast about the future. What matters is having analysis that can reveal gaps in our understanding of the situation.
>
> *(Kucharski 2020, 154)*

He cites a 2015 paper by Whitty:

> They are generally most useful when they identify impacts of policy decision which are not predictable by common sense. The key is usually not that they are "right," but that they provide an unpredicted insight.
>
> *(Whitty 2015, 4)*

In fact, neither this quote nor the paper from which it is drawn supports Kucharski's interpretation. In the passage from which the quote is lifted, Whitty is not commenting on models at all, but on their presentation and interpretation. The passage leading up to the sentences just quoted reads as follows.

> Models unfortunately tend to induce extreme reactions in most people, including policymakers, who are not modelers. Either they tend to believe them completely, or ignore them completely. Clearly the right place is between the two. The tendency of some modelers to present them as scientific predictions of the future rather than models does not help. Models are widely used in government, and some models arguably have too much influence.
>
> *(Whitty 2015, 4)*

All of this could apply equally to the interpretation of a mechanistic model, including the understanding it offers and its counterfactual consequences, as well as to a descriptive model making a projection about the future. The message is not:

"Model the impact of policy decisions," nor is it "Model underlying dynamics so we understand what is going on." It is "Present and interpret models with an appropriate degree of scepticism, and neither seek for them nor accord to them excessive or inadequate influence." Whitty's paper could be quoted against ICL9 for its huge influence, much more readily than in support of its modelling approach.[2]

Whitty's paper is written from the perspective of the then-recent Chief Science Advisor of a government department[3] on what makes academic papers useful. He discusses a range of types of papers, modelling being one. Many of the features that Whitty emphasizes as making models useful are violated by the products of ICL, notably the feature that policy-makers should have an interface through which they can modify assumptions they disagree with, and that papers should be as simple as possible, synthesize the breadth of scientific opinion, and involve economic analysis. The one that Kucharski identifies – greater understanding – is not, in fact, on Whitty's list at all. The fact that Kucharski reads Whitty in this way is therefore informative as to the intellectual climate within epidemiology, and especially within infectious disease modelling.

More likely, this desire for understanding connects with the ability of models to identify counterfactual outcomes and thus predict the outcomes of interventions. Of course, policy-makers do want to know about the outcomes of interventions. But the desire of *modellers* to predict the outcomes of interventions is a different thing. It is not necessarily what policy-makers want, or at any rate what they need, from an individual model. On the input side, more factors may influence the outcome of interest than a model can possibly model, since there may be qualitative factors that just cannot be quantified or otherwise treated as a variable. On the output side, there are almost certainly outcomes that a model will not include, notably economic and political ones (where the model is an infectious disease model). The desire to model the effect of policy interventions has a source in epidemiological intellectual ambition, perhaps the same one that leads advocates of counterfactualist approaches to causal inference to question the point of estimating causal effects without identifying an intervention. And, even if projections about the outcomes of policy decisions seem to provide policy-makers with the information they need to make their decisions, no model can truly do that. Both the relevant inputs and the outcomes of interest to policy-makers will always outrun anything a model can represent.

2.4 The Influence of London

We will shortly turn to a report that influenced Pandemic strategy globally to an extent that no other modelling report ever has. This was the Imperial College Covid-19 Response Team Report 9. Henceforth, we will refer to this report as ICL9, and to this response team as ICL (which does not, therefore, refer to that entire university).

Before looking at the report itself, it is important to understand the institution from which it arose. ICL9 was produced by the Imperial College London Covid-19 Response Team at the WHO Collaborating Centre for Infectious Disease Modelling. The WHO has over 800 such affiliated centres globally (WHO 2024a). Three others include "modelling" in the name (WHO 2024c). Their topics are malaria, geospatial disease, and non-communicable disease. Of these, the latter is the most potentially relevant, and that centre is also hosted at ICL. This university, therefore, provides half of the WHO's dedicated disease-modelling capacity, and the only capacity dedicated to infectious disease modelling. The WHO follows a principle established in 1949 that prohibits it from establishing its own international research institutions and commits it to supporting and making use of existing activities of existing institutions (WHO 2024a). Thus, the position of a collaborating centre is a powerful one with the WHO, since the organization does not have its own independent scientific operations to generate the knowledge the centres provide – nor to check it.

It is not, therefore, just a cheap shot to point out that the word "Imperial" in the institution's name harks back to the former British Empire. After World War II, European empires fell away. However, many imperial power relations persisted, and still do persist, in many of the international organizations set up in that era. The WHO is one of these, and the role of ICL's infectious disease-modelling group is one rather specific instance of these relations. That group bears no special responsibility for this situation: the relationship is one among countless similar echoes of former imperial dynamics. But it is one that suddenly proved extremely important in 2020.

The connections between WHO and its collaborating centres are public, but they are not necessarily transparent. It is not clear in which direction power and influence flow, which individuals or offices are central, whose interests lie where, what agendas might be at stake, or what previous alliances or grievances exist. It is not possible to infer much from the existence of the visible connections. However, they are important to know about when one considers how plausible it might be that a particular action was influential, whether there was a degree of coordination, or whether stated reasons are accompanied by unstated motives.

The terms of reference for the ICL WHO Collaborating Centre for Infectious Diseases give it both political and epistemic powers. They are as follows:

1 Upon request of WHO, provide rapid analysis of urgent infectious disease problems, notably outbreaks and events of international concern.
2 Upon request of WHO, provide technical assistance to WHO infectious disease programmes, including coordination of expertise in modelling, and contribute to WHO information products.
3 Upon request of WHO, contribute to capacity-building in modelling in accordance with WHO needs and planning (WHO 2024b).

These terms of reference position ICL not only as a policy advisor but as a custodian of expertise, empowering ICL's modellers in relation to others and conferring epistemic power upon them. It is easy to see that 1 confers a kind of power, because it implies that the WHO will seek information from ICL and that the information ICL provides will have special significance for the WHO. But the function of 2 and 3 as distributors of power is less obvious, since the power is presented in terms of technical assistance, coordination of expertise, contribution to information products, and capacity-building. These terms sound neutral, benign, and even generous. But it is inescapable that an actor empowered as a technical assistant, coordinator of expertise, contributor to information products, or capacity-builder will thereby be empowered to define what is technically possible, necessary, or correct; what then counts as the expertise that needs coordinating; what the "information products" – presumably, reports and the like – say; and which capacity gets built, and where.

In short, ICL is an extremely powerful institution in the global infectious disease-modelling community; possibly, the most powerful. Most of the powers are not codified or formal. There is no rule that WHO must do what ICL says, or vice versa. Nor are the powers surreptitious. There is no secret society or underground network. These are not the uncanny coincidences that provide fodder for suspicions of conspiracy; they are institutional arrangements carrying little formal power but indicating significant informal ("soft") power. It is a system of continual negotiation, which is the norm in academia and diplomacy alike. Power exists in such relationships, and the institutional ties here render it plausible that ICL9 had the impact we will attribute to it in this and subsequent chapters.

2.5 ICL9

Against this background, the Imperial College Covid-19 Response Team was formed by this centre and started producing reports. From the outset, it "adopted the policy of immediately sharing findings on the developing pandemic," and published 51 reports on its website between 17 January 2020 and 14 February 2022 – 30 of them in the first six months (Imperial College Covid-19 Response Team 2020).

On 16 March, a team of 27 authors at ICL published on its website a report titled "Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand" (Ferguson et al. 2020), which we henceforth refer to as *ICL9*. Probably, ICL9 was the most influential mathematical model (or rather, report of a modelling result) ever. It influenced both policy and science. Although it does not actually model or explicitly recommend a full lockdown, it created a framework in which lockdown was justifiable (see

Chapter 4). In science, it influenced the methods that were used for predictive modelling during the Pandemic, and constrained the results that those models could produce. This influence is what we call methodological rigidity, and it is the subject of the present chapter.

ICL9 is based on a microsimulation model originally developed for influenza, and adapted to Covid-19 based on what was known at the time. This was necessary since it is an extremely complicated model, as microsimulation models are. The model seeks to simulate social contacts between members of a population and predict how the disease will spread on that basis. It might ascribe higher or lower rates of infection for different groups: for example, the model assumed that school children would infect each other at double the rate of adults because this is realistic for influenza. That is an example of an unfortunate maladaptation of the model to SARS-CoV-2, since schools were not a driver for the Covid-19 Pandemic. This, however, is not the kind of criticism we are interested in here, since it arises from a lack of information and is corrigible.

The report presents results of modelling the results of reducing social contact in a variety of ways, such as case isolation, school and university closures, no large gatherings, and so forth. The "do nothing" scenario is also modelled. This scenario is truly terrifying: 81% of the population in the United Kingdom and the United States would be infected, with 510,000 and 2.2 million deaths in the United Kingdom and the United States, respectively, even before considering the effects of the healthcare system being overwhelmed. The report includes a strategic distinction between suppression and mitigation, the subject of the next chapter, and, based on this distinction, concludes that the minimal viable level of intervention to achieve suppression would be "population-wide social distancing combined with home isolation of cases and school and university closure." The best possible "mitigation" strategy consists of "case isolation, household quarantine and social distancing of the elderly" and would roughly halve the do-nothing numbers, but even so, "we predict there would still be in the order of 250,000 deaths in GB, and 1.1–1.2 million in the US."

This report was a key factor in the United Kingdom's decision to lock down on 23 March 2020. Possibly, its public release was intended to encourage political action. It obviously represents the tip of a larger advisory exercise, presenting, as it says, "the results of epidemiological modelling which has informed policy-making in the UK and other countries." This indicates its global influence, as one would expect given ICL's role in the WHO. The "other countries" are not indicated, other than the United States, and the rationale for that is likewise not explained. The report provides less numerical detail but at points says it is "qualitatively similar." The question of its influence in the United States, however, is complicated, and it is to this we now turn.

2.6 The Institute for Health Metrics Evaluation

The IHME is an independent institute based at the University of Washington, where it is located within the Hans Rosling Centre for Population Health, made up of an international network of research scientists. The term "population health" usually indicates a concern with equity and with the health of particular underserved or disadvantaged *populations*, as opposed to the older and more general term *public* health. This Centre was established in 2016 with a large donation from the Bill and Melinda Gates Foundation. The IHME itself was created earlier, in 2007, also with support from that foundation and from the University of Washington (IHME 2024). It addresses global health issues, as one would expect from its association with the Gates Foundation (and with Hans Rosling, an advocate for health in low-income settings).

In 2018, the IHME signed a Memorandum of Understanding to collaborate to "improve health data globally," working together on the annual Global Burden of Disease study published by the IHME – the largest of its kind. It is hard to tell how active this collaboration has been; the WHO is not mentioned on the Global Burden of Disease web page, and the WHO appears to produce its own parallel report, without any obvious mention of the Global Burden of Disease report or the IHME. In other words, the connections between IHME and WHO are not clear from an external perspective, and they are certainly not as strong as those between WHO and ICL, nor do they suggest that IHME occupies a powerful position in relation to the WHO, or that the WHO might rely on IHME as a primary or sole source of expertise or information in relation to a given topic. The WHO seems to have its own similar operation, and there is nothing to indicate that the WHO will turn to the IHME for certain sorts of information, nor that the IHME expects to receive such requests or is expected to respond to them. Its position in the global health network is thus quite different from that of ICL.

It is also important that the IHME is explicitly a "global centre," and that its emphasis is on global health, while the WHO-ICL centre is very much an ICL entity, featuring prominently on ICL's website as an advertisement for the university's "impact," not presenting itself as a global collaboration; and it includes a strong methodological emphasis (specifically, mathematical), evident too in many of the 51 Covid-19 reports. A tension between global health-oriented epidemiologists and mathematically oriented epidemiologists was visible during the Pandemic, and it is striking how many of ICL's reports and publications (but certainly not all) have the feel of research papers presenting mathematical or methodological innovations. Whether the methods they present were useful or widely used is open to question.

2.7 The IHME Model

On 26 March, the IHME placed a paper on a preprint server presenting modelling results of healthcare demand in the United States.[4] It identifies itself, accurately so far as we can tell, as the first state-by-state analysis. This predicted a significant excess demand in the United States for total hospital beds and ICU beds in the second week of April, and peak demand varying by state in April and May (Murray 2020; Team and Murray 2020). On the same day, it was reported in a news release on the IHME's website with the take-home message in the headline rather than buried in the discussion: "US hospitals could be overwhelmed in the second week of April by demand for ICU beds, and US deaths could total 81,000 by July" (IHME 2020). Notwithstanding the strong headline, the byline was more positive: "Assuming Strong Social Distancing, This Wave of the Pandemic Could End in the US in June." The paper's recommendations are bullet-pointed in the news story as:

- Postponing elective procedures
- Increasing the number of beds above licensed capacity
- Establishing emergency field hospitals
- Reducing staff-to-patient ratios (IHME 2020)

The model is significantly different from ICL9 in both its methods and its outputs. Methodologically, it takes a curve-fitting approach, using data from elsewhere to construct an epidemic curve, and then seeing how this would play out in the US states. It is a descriptive approach. It does not, therefore, seek to model differences between the places where the data are from and the targets of the model, nor differences between different possible intervention strategies. It simply extrapolates from what has happened so far, based on what happened elsewhere.

The justification for this approach was that epidemic curves in different places were roughly the same, and so provided a good starting point. As more data came in, the fit could be improved, and the modellers updated their projections regularly with data from the states they were modelling. It was, in other words, an ongoing exercise, and one that resulted in an interface that was updated on an ongoing basis. On the other hand, ICL9 was a report that was published once, and the underlying model was not suitable for ongoing updating – indeed, daily data were not directly relevant to its projections, but would inform fresh estimates of underlying parameters.

The IHME paper differs in its projections. The scenario in which ICL9 predicts 1.1–1.2 million deaths is, roughly speaking, the scenario that was then in place in many US states. The IHME predicts 81,000 deaths in (roughly) the same scenario. This is still a big number, but it is obviously far less than what ICL9 predicts.

This might seem to be a significant difference. But there is a more significant one. ICL9 is an early report, and its accuracy is less important than its message: that major action is required immediately to prevent dire outcomes. IHME predicts dire outcomes, just not quite so dire. But it does not support immediate, major action in the same way.

This is because its methodology assumes that the *shape* of the curve is roughly similar in different places, independent of their intervention strategy. "We are becoming more confident that the shape of the curve [is accurately] informed by locations outside the U.S.," a professor at IHME told one reporter (Begley 2020). But the locations from which the model draws data are very diverse, from China to Italy, and they vary widely in their response strategies. This runs contrary to ICL9's central finding, which is not any particular number, but a stark difference between the shapes of the curve, dependent on intervention strategy.

IHME's model was not saying that interventions make no difference: they might make a major difference to mortality, but not to the shape that the mortality curve takes – only to the absolute levels it reaches. To be clear, the IHME report does *not* provide evidence against the effectiveness of NPIs, and its authors repeatedly advocated for NPIs. The IHME's model is not an anti-lockdown manifesto.

Rather, it contradicts the idea that there is a *sharp divergence* between different scenarios. This is very important for the lockdown strategy, much more important than the simple idea that NPIs make a difference, because it warrants all-or-nothing thinking. We discuss this relationship between methodology and strategy in the next chapter. The point for present purposes is that the dramatic effect of NPIs is the central conclusion of ICL9. But the idea of a sharp divergence between alternative scenarios is not borne out by the IHME model. This is obscured by the fact that IHME does not model hypothetical scenarios. But it is implied by the methodology, which assumes that interventions do not result in step-change differences, and therefore that curves can be transposed to the US states and then adjusted. The methodology, therefore, was the subject of an attack.

2.8 The Attack on the IHME Model

The story was prompted by a paper published in the *Annals of Internal Medicine* by three authors at the London School of Hygiene and Tropical Medicine, University of California, Berkeley, and ICL, titled, "Caution Warranted: Using the Institute for Health Metrics and Evaluation Model for Predicting the Course of the COVID-19 Pandemic" (Jewell, Lewnard, and Jewell 2020). In essence, this paper alleges that the IHME model is unscientific:

> The IHME projections are based not on transmission dynamics but on a statistical model with no epidemiologic basis (Jewell, Lewnard, and Jewell 2020, 1).

This is as close to saying that the IHME model is unscientific as one could expect to find in a published commentary by one group of scientists at another. It is an astonishingly blunt criticism. It is unclear what "epidemiologic basis" means, but presumably, it means simply that the model does not derive projections by modelling transmission dynamics but by identifying patterns in the data. To call this "no epidemiologic basis" is flagrant rhetoric because the types of models used in epidemiology are highly diverse. In science generally, the quality of a model is gauged by its ability to tell us something useful about the target system (Suárez 2024), and making good predictions is useful in itself. There is simply no requirement that good scientific models – epidemiological models included – must represent specific dynamics of the target system. The appropriateness of the model depends on the purpose for which it is used.

The model is summarized as extrapolating "similar patterns in mortality growth curves to forecast expected deaths," based on worldwide mortality data. In other words, the IHME model takes information about what has happened elsewhere and projects what would happen in the areas of interest if the epidemic curve has the same shape. This is a considerable simplification, but it captures the essence of the methodology. It also places the IHME squarely within the camp of descriptive or statistical models.

The paper then goes on to offer six criticisms. It is not our purpose in this chapter to refute them, evaluate ICL9, or defend IHME. However, sociologists of scientific knowledge, such as David Bloor, have pointed out that when the scientific arguments in a debate do not add up, this can be important evidence for thinking that the real issue lies elsewhere (Bloor 1991). With that in mind, we now briefly examine the six criticisms in the article before concluding.

2.9 The Six Criticisms

First, "The model rests on the likely incorrect assumption that effects of social distancing policies are the same everywhere" (Jewell, Lewnard, and Jewell 2020, 1). This is not the case. The IHME model simply assumes some similarity in the shape of epidemic curves for different regions. The IHME model captured the human behavioural response to the pandemic, while ICL9 assumed people would not change anything at all, according to Christopher Murray from IHME (Findl and Suárez 2021). The amplitude and size of waves could still differ from place to place for many reasons, including differences in the effects of social distancing policies. Moreover, the shape of the curve was being updated on a daily basis. Where differences in – for instance – the effect of social distancing policies existed, the projections would come to reflect these with growing accuracy as they were updated.

The critical paper also rolls into the first objection a second point: "Short term forecasts based on curve fitting methods have no ability to predict second waves" (2020, 1). But the purpose of short-term forecasts is not to predict long-term phenomena such as subsequent waves. The presentation of a single wave in a forecast should not be interpreted as there is only going to be a single wave. What is more, mechanistic models are no better at predicting second waves unless they include information about waning immunity or new variants, and at the time, models could either assume permanent immunity or hazard a guess about how immunity might wane or the virus evolve. There was no model, and there could have been no model, that could make informative predictions about second waves.

Second, "countries that have flattened death curves earliest may not provide a basis for extrapolating trends in areas where similar control could prove elusive" (2020, 1). This is correct, but a good descriptive approach would take account of this and would use comparable examples to inform forecasts for their own region. Using South Korea as a guide to Italy is less useful than using Italy as a guide to Belgium. However, IHME found that they were able to use information from some countries as a guide to what would happen in others, notwithstanding differences of various kinds, including policy response. It was built into their approach to detect inaccuracy since the model was updated daily. Indeed, the IHME's approach would detect whether control was "elusive" in a given country as the wave developed, which in turn would be useful information to inform policy. There is no valid criticism here.

"Third, death counts can be unreliable and reporting differences occur even within regions" (2020, 1). Reporting delays and undercounting do happen, and should be taken into account. However, to know what is really happening, you have to work with the data at hand and do the best you can to work around the data problems. Curve-fitting approaches can help in the discovery of major reporting problems or other important changes, such as the emergence of a new variant, if the data from one region depart from an expected pattern. Data problems do not automatically warrant ignoring data and working solely from first principles. Conversely, using data that have problems is not wrong, provided caution is exercised.

Fourth, the paper criticizes IHME for showing a precise central line within their prediction envelope (2020, 1). However, depending on the purpose of a short-term forecast, there is nothing wrong with an indication of the most likely outcome. On the contrary, it is useful to know what the modellers consider most likely to happen, and not merely that it will fall anywhere between two bounds. Identifying a most likely outcome also prompts modellers to investigate retrospectively why they were wrong (where they were) and improve their models. Moreover, if a most-likely outcome is included, this may give an early warning that there is a departure from an existing pattern, for example, a new variant in circulation. False precision is obviously bad, but indicating a most likely outcome – along with appropriate indications of uncertainty – is not bad, and often good.

"Fifth, updated projections already reveal substantial volatility. For New York, the model predicted 10 243 deaths (range, 5167 to 26 444) on 27 March and 15

546 (range, 8016 to 22 255) on 30 March" (2020, 1–2). This criticism is unwarranted. The updated forecast remains within the prior forecast range, with both the lower bound increased from 5,167 to 8,016 and the upper bound decreased from 26,444 to 22,255. In other words, these figures show that performance is improving over time. These numbers were emphasized in the media, which is notoriously hostile to changes of mind. Yet these adjustments were not an indication of modellers changing their mind, merely updating in response to new data. In prediction, updating your beliefs in light of new evidence is rational and necessary, and the ability to do so is identified as a key indicator of predictive success by those statisticians and psychologists who have made a study of it (Tetlock 2005; Silver 2012).

> Finally, the projections are being interpreted misleadingly in formal and social media, without sufficient caveats, and outcomes differ substantially from those of other models. Upper uncertainty bounds are being interpreted as "worst-case" when, at best, they reflect only one scenario. The model yields an attack rate less than 5%; higher rates will lead to greater mortality than upper bounds (Jewell, Lewnard, and Jewell 2020, 2).

The purpose of short-term forecasts is not to produce counterfactual scenarios, but to provide a forecast as to what will most likely happen if a current pattern continues. If the forecast points to an attack rate of less than 5%, then the only relevant question is whether the attack rate ends up at less than 5% at the end of the wave. The existence of other counterfactual models with different outcomes is no reason not to produce accurate forecasts based on what is actually happening.

If a current pattern continues unchanged, or if it changes significantly after the implementation or lifting of specific restrictions, then it tells us something about the potential effect of implementing or lifting restrictions. This information can be used to evaluate and calibrate other mechanistic models that produced counterfactual scenarios for different policies. Statistical curve fitting approaches can therefore complement mechanistic approaches to modelling and vice versa.

The paper concludes with a more reasonable paragraph that is, however, contradictory to its belligerent opening stance and the six criticisms. It says that the "IHME's model may be reliable only for short-term projections." Notwithstanding "only," this contradicts the strong opening claim that the model has no epidemiological basis, since a model with no epidemiological basis would presumably not be useful for any epidemiological purpose, including short-term projections. It also remarks that "It is also unlikely that a 'one-size' model will fit all regions at all times" and that "Policymakers will be best served when they consider projections from multiple models." Again, these remarks appear to make space for a use for the IHME model, albeit implicitly. However, from a rhetorical perspective, the generalities of this final paragraph do nothing to rehabilitate this particular model. The implication of the piece as a whole is that the "multiple models" referred to do not include that one, and that the IHME model, *and its methods,* are best not used at all.

The six aforementioned criticisms are intended to support the central claim that the IHME model deploys a method without an epidemiological basis. However, none of the criticisms of the methods used by IHME is borne out by actual scientific practice. One might say that none is scientific. Infectious disease epidemiologists, in fact, use descriptive approaches extensively for purposes of short-term forecasts, which is what IHME was producing. That is because they are useful for short-term forecasts.

This point is made by some of those involved in the IHME model in a 2021 paper, which may well be a veiled response (Friedman et al. 2021). The point is borne out in a 2022 paper in which a different team evaluated 136 modelling papers and showed that most modellers were using statistical methods for short-term forecasting during the Pandemic, especially later on (Nixon et al. 2022). Findl and Suárez (2021) further pointed out that descriptive models such as the IHME model provide what they call "descriptive understanding," which is essential in the process of gaining "explanatory understanding." In Chapter 3, we will discuss the South African experience, where a switch from mechanistic to descriptive modelling improved forecasting performance. In short, there is no basis to the central claim: descriptive modelling has an epidemiological basis, namely, its utility for short-term forecasts.

Actually, descriptive modelling has a much wider utility than just short-term forecasting. In the next chapter, we show how it could have been used for calibrating and retrospectively evaluating mechanistic models. Johannes Findl and Javier Suarez give a detailed account of the uses of descriptive models such as the IHMEs, and argue that there are different kinds of understanding, some of which descriptive models are best suited to provide (Findl and Suárez 2021). In hindsight, to say that the IHME's projections had "no epidemiologic basis" was itself a claim without an epidemiological basis.

2.10 Media Coverage

It is safe to say that few journalists make a habit of reading the *Annals of Internal Medicine.* Yet following the publication of the Jewell paper, the IHME model was attacked in the media.

The effect of that attack can be seen by the reversal in the tone of coverage. The IHME report was initially covered by *Fox News* (which is conservative) on 7 April 2020, in a piece titled, "New York has been crushed by COVID–19. Will other US cities be spared?" (Hickok 2020). The answer suggested by the piece was – maybe: "It may not be too late for mitigation measures to help cities avoid a health-care system collapse, if the measures are taken seriously and strictly adhered to." There was not much of a media splash, and what coverage there was certainly did not take the model as licensing complacency.

However, following the publication of the Jewell paper a few weeks later, a number of media outlets attacked the model. These were liberal media outlets, of a kind probably inclined to consider themselves "pro science." And they quoted scientists. "It's not a model that most of us in the infectious disease epidemiology field think is well suited," said one; "That it is being used for policy decisions and its results interpreted wrongly is a travesty unfolding before our eyes," according to another; and a third called it "highly inaccurate" (Piper 2020; Begley 2020). The stories were accompanied by huge pictures of Donald Trump's glowering face, an image obviously likely to cause a negative emotional reaction in the target readership of this sort of publication – a readership likely to include peers, funders, and employers of the IHME model's authors.

Yet in retrospect, the IHME model was not only defensible – it was, in some respects, superior. It provided state-level analysis when ICL9 did not; it was more easily updated; and it offered a more pragmatic interface for policy-makers. Crucially, it turned out to be more accurate in some key respects. ICL9, for instance, significantly overestimated the time to peak in the UK, an error that may have delayed action. This error arose from methodological commitments that the IHME's approach, by design, avoided. However the IHME model presented a methodological and institutional threat to ICL. The campaign against it, scientific and public, should be understood not simply as a dispute about best practices, but as a struggle over what kind of science would be allowed to guide policy.[5]

2.11 Conclusion

We are all accustomed to the idea of a politically motivated scientific attack. More generally, we are used to the idea that scientific theories, findings, postures, and so forth can be motivated by political commitments. We would expect a climate change paper funded by certain groups to reach certain results, for example.

But this episode is an instance of something different. The core issue here was not about politics in the international, national sense, but about politics in and around science: about who would wield the influence suddenly available to scientists at that time. This was not a case of politics corrupting science, but of scientific institutions using political strategies to consolidate epistemic authority.

The modelling methods used by the IHME model contradicted the core result of ICL9: that interventions would bring about step changes in the shape of the epidemic curve. This conclusion was the basis of the political advice being given by the scientific establishment to political leaders. ICL is the scientific establishment in this context due to its unique role in the WHO, while the IHME is the outsider.

The IHME's model provided the basis for a different type of policy, not because its predictions were more moderate (since they were still serious), but because its *methods* failed to support the core idea: that a sharp distinction existed between different intervention scenarios. It used a different method from that of ICL, and ICL was in a position to discredit that method and insist on the superiority of its own approach.

In retrospect, the IHME was vindicated in its methodological decisions and unfairly treated as regards accuracy, certainty, and politics. The episode exemplifies the phenomenon explored in the next chapter: *methodological rigidity*.

Notes

1 Johannes Findl and Javier Suárez argue that while mechanistic models may yield "explanatory understanding," descriptive models yield "descriptive understanding," which is also important (Findl and Suárez 2021). If correct, this would further show that the explanatory potential is not a decisive mark of a model's superiority.
2 Not to imply that Kucharski used it this way, of course.
3 The Department for International Development (DFID).
4 The timeline is a little confusing. The preprint is reported on the website on 26 March but dated 30 March on the preprint server. We assume an update although we find no record of one.
5 A different defence was advanced by Drew Schroeder in 2021, who argued that the IHME's predictions had been wrongly interpreted

3

METHODOLOGICAL RIGIDITY

3.1 Introduction

The previous chapter focused on a particular episode, but an episode of what, exactly? The present chapter proposes an answer to that question. What played out between the London-based modelers and those in Washington was an instance of a wider *methodological rigidity.* In this chapter, we explain what we mean by that term, why we think it occurred in Pandemic-era epidemiological modelling, and why that was a bad thing.[1]

There is a large philosophical literature on models in science, addressing such questions as what they are, whether they represent, what different functions they perform, how they relate to other scientific entities (especially theories), and, recently, how non-epistemic values can shape their construction and output (see inter alia: Godfrey-Smith 2006; Parker and Winsberg 2018; Frigg 2022; Knuuttila, Carrillo, and Koskinen 2024; Winsberg and Harvard 2024; Frigg and Hartmann 2025). This literature has focused on a few sciences, particularly physics, economics, biology, and climate science – but not epidemiology. When the Pandemic emerged, this changed, and some philosophers of science actively debated the performance and role of Covid-19 models. For example, Eric Winsberg and colleagues argued early in the pandemic that model-based projections were too uncertain and value-laden to justify drastic policies (Winsberg, Brennan, and Surprenant 2020). Critics responded that even imperfect models can guide urgent decisions when used judiciously: van Basshuysen and White contend that Winsberg et al.'s case against lockdowns undervalued the available evidence and context (Basshuysen and White 2021; White, van Basshuysen, and Frisch 2022). Robert Northcott argued that some models were better than others:

DOI: 10.4324/9781003732495-4

overly ambitious models faltered by assuming uniform stability across contexts, whereas more modest, context-tuned models proved more effective (Northcott 2022). As mentioned in the previous chapter, Suárez and Findl argued that descriptive approaches were being unfairly discredited by some powerful modellers (Findl and Suárez 2021). This is a mere selection: there were many more.

Yet none of this literature fully makes sense of what happened in the episode described in the previous chapter. Nor is there anything in the wider modelling literature that one might point to and say, "Ah, this is an instance of that thing philosophers have talked about for so long!" Our goal in the present chapter is to fill that gap.

The reason there is a gap is that the problem is not really about models per se at all, but about methods more generally. The problem was deciding which method was the best one to use in the circumstances. It so happened that the methods concerned were modelling methods, but where similar tensions have arisen in the past, they have concerned other things, particularly evidence and causal inference. Most famously, the term "evidence" has been redefined by the evidence-based medicine and evidence-based policy movements, so as to prioritize certain kinds of information over others – especially, prioritizing randomized trials over observational evidence, mechanistic reasoning, or expert opinion (Straus et al. 2018). A similarly-inspired development in observational epidemiology led to a strict and formal definition of the term "causal inference" in place of the older, explicitly pluralistic approaches (Robins, Hernán, and Brumback 2000).

Philosophers have commented extensively on evidence-based medicine (Howick 2011; Holman and Sung 2025) and policy (Cartwright and Hardie 2012), and somewhat on causal inference (Broadbent 2021), and in most cases have tended to push back against the rigidifying approaches. However, these discussions have focused on causation. Thus, they, too, are not directly applicable to the situation that we are trying to understand in the Pandemic, when epidemiological modellers started dismissing each other's methods for *predictive* modelling.

Yet despite the fact that the ultimate point of using evidence in policy is to support prediction, philosophers of evidence have curiously neglected the predictive dimension itself. They have instead tended to focus strongly on whether and how causation can be inferred, and if it can, whether there is a good inference available that it will hold in a contemplated situation. This way of seeing the problem simply does not resemble the way epidemiological modellers think, whether they are mechanistic or descriptive in their approach. For instance, the currently popular idea that you need both evidence about an association and evidence about a mechanism in order to make a causal inference (Shan and Williamson 2023) simply gets no purchase on the dispute between ICL and IHME, since they were not talking about associations or mechanisms. Perhaps a devotee could shoehorn that episode into the "Evidential Pluralism" framework, but a critic would doubtless shoehorn it straight out again, and it is not clear what would be gained by the ensuing discussion. Our goal in this book is not to use the Pandemic to do philosophy, but to use philosophy to understand the Pandemic.

In the present chapter, then, we set out our own account of what we think went wrong: *methodological rigidity*. Methodological rigidity, as we argue here, is one key mechanism through which scientific orthodoxy operated during the pandemic. Having defined methodological rigidity, we spend some time setting out what makes for a good prediction and distinguishing between different kinds of prediction. We then explore two predictive errors made by the UK's Scientific Advisory Group for Emergencies (SAGE) in 2021. These errors are picked because they are clearly errors, and because they occurred well over a year into the Pandemic. We contrast them with better performances by other teams – South African and Swedish – and argue that the difference was due to the fixation of the teams driving SAGE's outputs on certain methods for modelling – mechanistic approaches, at the expense of others. We finish with some suggestions about how to avoid methodological rigidity. These might not seem Earth-shattering, but we argue, they could have been life-saving.

A note on terminology before proceeding. Elsewhere, we have explicitly endorsed "methodological pluralism" (Vandenbroucke, Broadbent, and Pearce 2016; Broadbent, Vandenbroucke, and Pearce 2016; Streicher, Broadbent, and Hellewell 2025). We have not defined this position, and have intended it as a very minimal one – a rejection of formalizing or rigidifying approaches. Now, with the concept of methodological rigidity in hand, we can define our minimal methodological pluralism as simply the rejection of methodological rigidity.

However, we do not frame our argument in terms of pluralism, defined this way or any other. The term invites philosophical debate. Some might regard our definition as too weak to deserve its own term at all, and prefer to reserve "pluralism" for more substantive positions. Others might think we mean the same thing as other authors who have a more restrictive notion of pluralism (Shan and Williamson 2023). All of this is beside our point, which, as we have said before, is to refine philosophical tools so as to understand the Pandemic, and not use the Pandemic as a backdrop for comparing our favourite tools. We therefore avoid the term "pluralism" in this chapter. We do not care about pluralism, and the philosophical questions it raises: we care about methodological rigidity, and the real-world harms it causes.

3.2 Methodological Rigidity and Pandemic Prediction

Methodological rigidity is

(a) Reliance on one methodological approach without due consideration of
(b) Its applicability,

(c) Its vulnerabilities, or
(d) Its track record, or of
(e) Other relevant approaches.

The condition is disjunctive, so not all these conditions need to be satisfied for methodological rigidity to be present, and there is no need to show that each is satisfied in the cases we consider.

The keywords in this definition are "without due consideration." These words make methodological rigidity a bad thing because of the normative force of "due." Sticking to a method if it has proven itself time and again may be perfectly reasonable. Methodological rigidity arises when there is good reason to doubt, check, evaluate, or otherwise reconsider, but instead the method is rigidly adhered to.

Why is methodological rigidity a problem? Because it inhibits finding out about facts generally, and in the context of predictions in the Pandemic, because it resulted in bad ones. In this chapter, we show that methodological rigidity was responsible for some bad predictions. We also show how some teams made better predictions, both early and late in the Pandemic, and despite being less well-resourced. The only explanation for the repeated failures of well-resourced teams, we argue, is their rigid adherence to a particular methodological approach.

3.3 What Is a Good Prediction?

The term "prediction" is used in a variety of ways, often confusingly. Here, we set out a framework for talking clearly about prediction. We say that a good prediction has three components: appropriate accuracy, appropriate precision, and appropriate method(s), and offer a definition on this basis. The notion of appropriateness is contextual and depends on the purpose for which a prediction is made.

3.3.1 Purpose

Predictions are not good or bad in themselves, but relative to the purpose for which they were made. This is the essential starting point for any attempt to evaluate a prediction.

An obvious example is the difference between predicting for the purpose of gaining knowledge of the future and predicting for the purpose of gaining knowledge about the theory that produces the prediction. The former is commonly what is meant by "forecasting." The latter is the kind of prediction that occurs in experimental science. Galileo predicted that a cannonball and a feather would fall at the same rate in a vacuum. The point of this prediction was not to enable us to make appropriate preparations for their arrival at the foot of the leaning tower of Pisa. The prediction is not even intended to approximate that situation. That tower is not in a vacuum, and applying Galileo's prediction in that context could prove either frustrating or disastrous. The point of the prediction is to demonstrate the truth of the theory under experimental conditions. If those conditions can be brought about by an experimenter, then the theory can be *tested* by observing *whether* the prediction is accurate.

On the other hand, if we use Newtonian mechanics to predict the trajectory of a satellite re-entering Earth's atmosphere, we are relying on the truth of that theory in an effort to find out about the future, perhaps so that we know where to send a recovery ship. We do not make this prediction in order to test Newtonian mechanics by sending a ship out to observe whether the satellite is where the theory predicts it should be. The context of testing differs from – in fact, mirrors – the context of forecasting, since in the former case we seek information from the event predicted, and in the latter we seek information from the theory or model that predicts.[2]

In discussions of modelling, a distinction is sometimes drawn between projecting and forecasting. This is sometimes an attempt to reflect the points just discussed. However, it is sometimes an attempt to defend poor predictions when the course of events reveals their inadequacy. A forecast is generally understood as a claim about the future and a projection, as a claim about what would happen under certain assumptions, without any assertion as to whether those assumptions obtain. The trouble is that inaccurate claims about the future are sometimes defended as "mere projections." But this is disingenuous, since it allows the modeller to escape the question of *why* they made these projections. Were they testing their model? Surely not, if the results are published in an advisory report. Were they seeking to inform policy? In which case, why did they proceed with unrealistic assumptions? Modelling to inform policy must provide information about realistic situations, not idealized ones.

The issue is that the projection/forecasting distinction purports to distinguish two types of activity but obscures the importance of purpose. The very same activity could be a forecast in one situation and a "projection" in another. A claim about what will happen if – say – a lockdown is implemented will be a projection, while the modeller does not know if there will be a lockdown, but a forecast if there is a lockdown, while remaining a projection if there is not. For this reason, Jonathan Fuller suggests that "the forecast/projection distinction is best invoked at the stage when model-users use a model's outputs to make predictions, to which those users may or may not attach an antecedent condition" (Fuller 2021, 46). What determines whether a certain modelling exercise is a "forecast" or "merely a projection" is not the kind of model used, or anything else about the mathematical activities of the modeller – nor even, strictly speaking, about the intentions or expectations of the modeller. It has to do with something extrinsic to the entire modelling process, which is whether the antecedents of the model are satisfied. A theoretician like Galileo might build a model to make a projection about the behaviour of feathers in vacuums, but an experimenter who succeeds in constructing an appropriate apparatus may use exactly the same model to make a forecast. As Fuller puts it, "a model does not know whether its assumptions obtain" (Fuller 2021, 46). There is no stable difference between projection and forecast, at least not one that is useful for evaluating the success of predictions. For evaluating predictions, what matters is their fitness for purpose – that is, their suitability for the use to which they are put.

3.3.2 Appropriate Accuracy

Accuracy is not an all-or-nothing matter. If a forecast is said to be "inaccurate," this could mean that it is out by 1% of some range, or by over 100%. The latter is surely sufficient to render the forecast pointless, but whether the former matters depends entirely on context. If we are talking about the location at which a satellite will land, and 1% relates to the trajectory, then the ship might end up hundreds of miles from the satellite. If it refers to a difference in daily hospitalizations during the Covid-19 Pandemic, on the other hand, a prediction 1% away from fact would be regarded as accurate.

A good prediction must therefore be one that is *appropriately* accurate.

3.3.3 Appropriate Precision

A prediction may be so vague as to all but guarantee its truth. To capture this, a distinction is commonly drawn between accuracy and precision. Accuracy refers to whether the prediction includes the facts within its scope. Precision refers to the scope.

Purpose is obviously relevant to precision, as to accuracy. In predicting the location of a satellite's landing, one need not arrive at the nearest centimetre or even the nearest 100 metres. Either way, the recovery vessel will sail to the same spot. So there is no point attempting the possibly impossible task of achieving accuracy to the nearest centimetre. On the other hand, in predicting likely peak hospital bed capacity, a prediction that demand will fall between 100 and 10,000 is useless for planning purposes. As we shall see below, a prediction of this kind was in fact made during the Pandemic. In this situation, whether the prediction comes true is really irrelevant since it leaves one none the wiser about what to do.

3.3.4 Appropriate Method

In the context of the Pandemic, some predictions were made for short-range planning purposes, some for the longer term. Different models were differently useful for these purposes (specifically, descriptive models were better for short-term and mechanistic models for longer term). Some predictions did not concern the future at all, but the present, for example, by seeking to estimate the number of infections on the basis of the number of detected cases. This kind of exercise is often known as “nowcasting.” (Again, this is primarily a matter of descriptive modelling.) There was also, of course, a distinction between planning and decision-making. A planner does not need to know what would happen, nor what would have happened – only what will happen; but a policy-maker needs to know what would happen if they took a decision they might not take, while an electorate may well wish to know what would have happened if a politician had done something different. Again, different modelling approaches are often appropriate for these different tasks.

In assessing whether a prediction is good, it is therefore important to know whether it was made with an *appropriate method* for the purpose at hand. This does not imply that one method is appropriate for one purpose, another for another. Often a combination of approaches will be appropriate – this is part of the message of the present chapter. The point, rather, is that the selection of methods needs to be evaluated in light of the purpose of the exercise.

3.4 Counterfactual Prediction

How should we assess a prediction based on circumstances that do not in fact obtain? For example, suppose a policy option was not selected because a model predicted it would have dire consequences, as was the case with ICL9 in particular. Alternatively, a prediction might have said that consequences of inaction would be acceptable, but the model might have been ignored, as was perhaps the case to some extent with IHME’s model.

In such situations, precision at least can be evaluated without any regard for the eventual outcome, since it is a function of the prediction and its purpose, and nothing else. Appropriateness of method may be evaluated to some extent, since there may be methodological critiques or over-reliance on a particular method. However, the appropriateness of the method may depend in part on the assessment of accuracy, since appropriateness of method concerns reliability, which concerns accuracy over multiple occasions – including this one. And where there are not many occasions to choose from, as in the Pandemic (especially early on), methodological evaluation may be inconclusive in the absence of information about accuracy.

Accuracy of a counterfactual prediction may be difficult to assess, but that does not mean that it is always impossible to find relevant evidence. One can sense-check and triangulate using alternative methods; evaluate the performance of a prediction retrospectively, as regards what it predicted about the policy that was actually implemented; present alternative models; and test the assumptions of the model. More generally, one can adopt a practice of always presenting the results of a descriptive model of a do-nothing scenario alongside mechanistic models of the various options. We discuss each of these proposals later in Section 3.13 where we discuss antidotes to methodological rigidity. The point for present purposes is that it is not necessary to give up on assessing counterfactual predictions merely because they are counterfactual.

3.5 Critical Care Bed Capacity in ICL9

In this section, we will use basic sense checking to evaluate the critical care bed estimates of the ICL9 "do nothing" and "optimal mitigation" counterfactual scenarios. ICL9 included a range of counterfactual mitigation scenarios, with response measures varying from "do-nothing" to "optimal mitigation," and concluded that "emergency surge capacity limits of the UK and US healthcare systems (will be) exceeded many times over" even under "optimal mitigation." Since the UK opted for suppression and eventually locked down, all the mitigation scenarios in ICL9 can be regarded as counterfactual scenarios. (US mitigation scenarios are not presented in the report.)

Sense checks can be done by comparing the actual relationships between key variables to those of ICL9, or by comparing model outcomes to real-world outcomes in countries that applied lighter restrictions or countries/cities that reacted very late.

At the peak of the ICL9 "do nothing" scenario, deaths were at 21 per 100k per day (Figure 1 – ICL9), and critical care beds were at 280 per 100k, a ratio of 1 death per day:13 critical care beds occupied (see Figure 2 – ICL9). However, in Bergamo, this same ratio never exceeded 1:2 throughout March 2020 (Polver et al. 2021), and in the UK, it never exceeded 1:3 (OWID data). The implication is that even if the deaths were correct for the ICL9 "do nothing" scenario, critical care bed numbers were inflated by at least fourfold.

Doing a similar comparison between the ratio of those hospitalized and those in critical care beds, ICL9 used a ratio of 30% (page 5, ICL9), while data at the time from Italy on 13 March 2020 pointed to a ratio of ICU beds occupied to hospital beds occupied of 9–11% (Remuzzi and Remuzzi 2020). Although not exactly the same metric this points to ICL9 critical care bed numbers being inflated significantly.

3.6 Attack Rate in ICL9

In this section, we apply similar methods as in the previous section to sense-check attack rate estimates.

Attack rate is the proportion of a clearly defined population that develops a given disease (or other health event) during a short, self-contained risk period. Although called a rate, it is really a cumulative incidence (a proportion): the time dimension is implicit in the outbreak window rather than explicitly in the denominator.

$$\text{Attack rate} = \frac{\text{Number of new cases during the outbreak}}{\text{Number of people at risk at the start of the outbreak}} \times 100\%$$

Attack rate is used in outbreak investigations because it is a simple, intuitive measure of how aggressive the outbreak is. An attack rate of 80% in a three-day window after consumption of some dodgy oysters at a wedding provides a good intuitive indication of an explosive outbreak, even if it is not, technically speaking, a rate, as the time period is not written into the formula. The denominator should include only those individuals who were genuinely at risk (for instance, attendees at the wedding where food poisoning occurred), excluding anyone known to be immune or unexposed (vegetarians, people allergic to oysters). Variants such as the secondary attack rate focus on cases arising from person-to-person spread among contacts of primary cases, helping to quantify transmissibility within households or other closed settings.

In the Pandemic, attack rate provided a simple, intuitive measure of how “bad” a wave was, or was expected to be. There was an important distinction between infections and detected cases, but unless otherwise stated, “attack rate” refers to infections rather than detected cases.

ICL9 estimated an (infection) attack rate of 81% for the “do nothing” scenario and an attack rate of 40% for “optimal mitigation.” The best sense check on these figures would be to consider the highest attack rates estimated at the time from real-world data. Attack rates can be estimated using deaths and the ICL9 IFR estimate of 1%. As of 20 February, China had reported 2,114 Covid-19 deaths. Most of these deaths were in Wuhan, a city of 11 million people, where deaths had already peaked. This pointed to an attack rate of 2% in Wuhan at the time, and an attack rate of around 5% by the end of the first wave in Wuhan. By mid-March, cumulative Covid-19 deaths were approaching 1,500 in Bergamo, Italy, a city with a population of 1.1 million people. Daily deaths had reached a peak by then. The attack rate at the time could be estimated at roughly 14%, pointing to an attack rate of 29% by the end of the first wave in Bergamo. An analysis published later confirmed this, finding an attack rate of 28.8%. However, actual deaths and actual IFRs turned out to be double that of initial estimates (6,028 deaths at an average IFR of 1.9%) (Polver et al. 2021). With the benefit of hindsight, no other city exceeded this level (29% attack rate) in all of Europe after the first wave.

In effect, Bergamo showed what could happen if Covid-19 was essentially unchecked in a Western European city with an unusually large proportion of elderly residents (31% of whom were over 60). It was essentially a do-nothing scenario, since action was not taken until much too late to influence the peak.

Attack rates are higher in cities (being densely populated), and lower when averaged across a whole country. An attack rate of 40% for the ICL *optimal mitigation scenario* at a *national* level was a third higher than the highest *city* attack rate (29%) that had played out under a *do nothing scenario*, which is to say, it was implausible.

These two errors combined – the critical care bed error and the attack rate error – amount to more than an order of magnitude, and partly explain why the "worst case" critical care bed estimates by the Swedish public health authority (discussed in Section 3.11 and shown in Figure 3.4) came in seven times lower compared to the "optimal mitigation" critical care bed estimates shown in ICL9's second figure.

Inflated numbers for counterfactual predictions obviously do not serve rational decision-making. But they create a further problem when it comes to evaluating the benefits of intervention measures. If the worst does not happen, then whatever was done will take the credit. The benefit will be overestimated, and costly measures – lockdowns – are more likely to be repeated, on the basis of erroneous risk–benefit evaluations. Moreover, modellers will be less able to predict the outcomes when lighter restrictions are introduced, as was inevitably the case almost everywhere, and this resulted in errors, as we shall see.

3.7 Timing in ICL9

In this section, we identify an important error concerning the timing of the start of the wave as well as the timing of the peak of the wave. ICL9 Figures 1 and 2 showed a later start and later peak of the first wave in the UK compared to what actually happened. Based on the SAGE meeting minutes before 13 March 2020 (the meeting at which the penny appears to have dropped), SAGE expected the UK to be 4–5 weeks behind Italy. This expectation of SAGE corresponds with ICL9. Given the institutional position of the ICL Covid-19 Response Team and the remark in ICL9 that it concerns modelling provided to the government, it is likely this model or a very similar model was discussed by SAGE before ICL9 was published on 16 March 2020.[3]

However, SAGE and ICL9 were both mistaken. The UK was much further along the epidemic wave than either of them realized. This error was pointed out on 12 March 2020 by Tomas Pueyo in a *Channel 4* interview that included Professor John Edmonds from SAGE. Pueyo was not an epidemiologist or even a scientist, but variously a blogger, expert on storytelling, and developer of an astrology app. Pueyo stated that the UK was only 13 days behind Italy (Law 2020). His argument was based on comparing the Italian case curve to that of the UK, the method that ICL and others were to criticize the IHME for using.

Pueyo's intervention has been called "stunning," for good reason. Warren Pearce provides an illuminating analysis of this episode, telling the story of a person with no epidemiological credentials, writing on an open blogging platform with no peer review process. Pueyo had previously blogged, for example, about what Skywalker can tell us about storytelling and why his astrology approach blew up on Facebook (Pearce 2023). He was on *Channel 4* on the strength of a blog post that received 40 million views in a week called "Coronavirus: Why You Must Act Now" (Pueyo 2020a), followed up by another whose title, "The Hammer and the Dance" (Pueyo 2020b), made it into German ministerial strategy (Heinen 2022). He was pitted against someone who was indisputably an expert, and represented SAGE's view at the time – and that of the British government – that drastic non-pharmaceutical interventions were not yet necessary.

Pueyo's stance was extremely normative, far more so, perhaps, than any scientist's would have been. The climax of his piece is almost a kind of poem:

> The coronavirus is coming to you.
> It's coming at an exponential speed: gradually, and then suddenly.
> It's a matter of days. Maybe a week or two.
> When it does, your healthcare system will be overwhelmed.
> Your fellow citizens will be treated in the hallways.
> Exhausted healthcare workers will break down. Some will die.
> They will have to decide which patient gets the oxygen and which one dies.
>
> The only way to prevent this [very bad outcome] is social distancing today. Not tomorrow. Today. That means keeping as many people home as possible, starting now.
>
> *(Pueyo 2020a)*

Of these factual claims, one provoked a sudden realization: that it was a matter of days, not weeks, before Covid-19 would be very widespread. In this respect, Pueyo was correct, notwithstanding his lack of relevant credentials.

The result was that SAGE realized with a shock that cases were already 30–40 times higher than they had previously believed. This realization was first reported in the 13 March 2020 SAGE minutes, immediately prompting an action to consider suppression.

How did an uncredentialed and highly opinionated individual in the blogosphere outperform the world's leading infectious disease modellers? By using methods that were more appropriate for short-term forecasting, and which focused on describing the actual data rather than simulating the epidemic wave from what was known of the underlying dynamics of disease transmission. Simulation models and compartment models are particularly poor when it comes to the temporal accuracy of events. This is not the purpose for which they were developed and generally used. The most accurate approach in this kind of situation would be to carefully scrutinize data from regions that are further ahead on the curve, using descriptive and statistical methods, and to compare this to local data emerging on a daily basis. Even though there would be differences between – say – Italy and the UK, the divergence becomes greater over time, meaning that short-term forecasting may be relatively reliable. Ironically, these were the methods used by ICL to estimate IFRs from limited data from China in January and February (Verity et al. 2020). ICL's estimate proved particularly accurate for older ages, where it counted (although it somewhat overestimated IFRs for younger ages). It was a good, useful estimate.[4] This IFR was relied upon in the mechanistic model used to generate ICL9, even though one might still object that China and the UK are very different and that, therefore, the IFR might be different. Yet ICL9 did not consider using contemporaneous data from other countries to generate a simple sense-check for its headline model predictions. It is as if the ICL team, and with them SAGE, were so focused on their models that they forgot about the data.

3.8 Two Errors by SAGE in 2021

We now turn to two high-profile episodes of poor predictive performance by SAGE and influential scientists. These are not the only errors of pandemic modelling, nor are these groups the only ones that made errors, of course. We pick them first because they are clear errors: what they predicted did not actually happen, even though the preconditions of the prediction were satisfied. Second, they emanate from institutions central to the Orthodoxy. Third, they illustrate methodological rigidity, which causes the errors. It should be obvious that we do not mean to condemn all predictions in the Pandemic, and indeed, we will contrast these cases with better predictions. But our argument does imply one general conclusion, extending well beyond these examples: that there is good reason for scepticism about all predictions infected with methodological rigidity.

The two episodes we focus on here occurred in July and December 2021. SAGE expressed high confidence about the prevalence and daily deaths respectively. On the first occasion, what actually transpired still fell within SAGE's confidence intervals, but they were so wide that this was not surprising, and accompanying statements are reasonably regarded as erroneous – in the sense that if anyone had tried to follow SAGE's advice, they would have been worse off. The second time, reality fell outside SAGE's very wide confidence intervals – a very clear error.

The first episode occurred in July 2021, when SAGE produced a report ahead of the planned easing of restrictions later that month (SAGE 2021a). The report predicted prevalence would "almost certainly remain extremely high for at least the rest of the summer." An influential member of the modelling subcommittee said in an interview that it was "almost inevitable" cases would rise to 100,000 a day (Ferguson 2021). In reality, cases had just peaked at 57,583 on 15 July 2021 (7-day average peak at 44,876) and continued to decline (UK Govt Coronavirus Data 2024).

The second episode occurred in December 2021, when it issued a consensus statement predicting between 600 and 6000 deaths daily in England if restrictions were not significantly tightened (SAGE 2021b). For the first time, the politicians entirely contradicted their scientific advisors, and were embarrassingly vindicated: daily deaths peaked at 202 (7-day average), three times lower than the lower bound of the range of outcomes modelled.

In addition to being erroneous, these reports included excessively wide confidence intervals. In July, SAGE predicted daily hospitalizations spanning 100–10,000: two orders of magnitude only one month after the lifting of restrictions. SAGE advised preparing for the upper limit. Being huge, the range of uncertainty encompassed the actual hospitalizations, which hovered below 1000 per day. But a guess might have done as well, and policy turns to science precisely to avoid over-preparing by 9000 beds. More embarrassingly, in predicting the effect of Omicron under light restrictions, SAGE again produced an excessively wide confidence interval (one order of magnitude), yet nonetheless overshot by a factor of 3–30.

The second of these episodes produced public furore. But public furore is not usually a good means of learning lessons. Accusations of a political or personal kind tend to obscure rather than illuminate the fundamentally important question: how could what is arguably the most powerful modelling community in the world get it so wrong? This is not rhetorical: it is a real and important question. Unless the question is answered, this kind of error will probably happen again, and until it is answered, we cannot begin to understand whether the policy decisions that were taken on the basis of its advice were justified.

3.9 Forgetting to Forecast

The most obvious way descriptive models can usefully accompany mechanistic models is by producing alternative contemporaneous forecasts. In December 2021, a new variant, Omicron, had arrived in the UK and was spreading in the UK. According to data from South Africa, it had already peaked there and was considerably milder than previous variants. It was therefore reasonably straightforward to fit a descriptive model to the available UK data, making adjustments for factors such as population age and prior immunity levels. At the time, Streicher's contemporaneous model produced a worst-case scenario of 350 deaths per day if the UK infection wave played out like Gauteng, South Africa's most densely populated province (Streicher 2021), or 295 per day for England on a simple proportion basis, where actual deaths peaked at 202 per day (Figure 3.1) – still substantially on the worse side of reality, as a reasonable worst-case scenario should

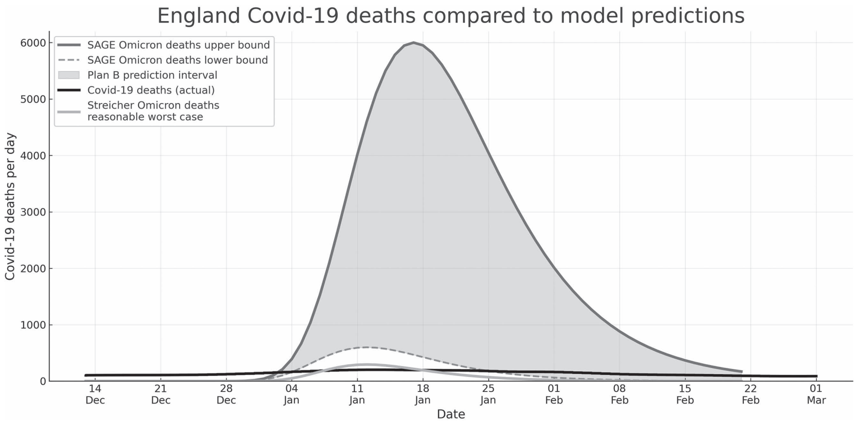


FIGURE 3.1 Projected and actual Omicron deaths in the UK

be, but considerably closer than SAGE's worst case.[5] The bank JP Morgan Chase also released a report with a much more conservative set of predictions, also using more descriptive methods and reasoning, which influenced the British government's decision to contradict its appointed advisors (Nelson 2022).

SAGE had this to say about South Africa:

> There are early indications for shorter lengths of [hospital] stay in South Africa [implying lower severity], however, it is unknown whether such observations would be seen in the UK, given differences in population demographics, COVID-19 epidemic timing and variant composition to date, vaccination types and programmes, health care systems, and so on (SAGE 2021b).

Instead, for its December 2021 forecast, SAGE based itself on two models from UK universities, both using similar mechanistic approaches. They did not consider any models of the simpler kind used by us, banks, or others.

SAGE is right to point out differences between South Africa and the UK. Yet the effect of these differences would have needed to be *enormous* to account for the difference between worst-case scenario of their mechanistic approach (6000 deaths per day) and that of a descriptive approach, such as Streicher used (a mere 295 deaths per day). The simple exercise described above can be done on a laptop by a postdoctoral researcher. If SAGE had done this exercise and published its results alongside the mechanistic modelling output, and *then* tried to explain why the South African experience was irrelevant, it would almost certainly have devoted more than one sentence to the matter. The explanation would need to be spectacular. More likely, it would have reconsidered both models, sought further information in an effort to understand why they were producing such different outcomes, and thought more deeply about whether South Africa really was so very different in the crucial respects for this prediction. Ultimately, it would probably have modified its results.

The ICL Response Team appears to have become aware that something was wrong with the estimates. Report 49, published on 16 December 2021, indicates that there was little change in severity between Omicron and Delta, but that the vaccinated were more prone to Omicron infections (compared to Delta), while those with a prior SARS-CoV-2 infection were 5.41 times more prone to reinfection (compared to Delta). In other words, Omicron was just as bad as Delta, more infectious, and would evade immunity from either vaccination or prior infection: terrible news. Six days later, on 22 December, Report 50 pointed to reduced severity, with a 40–45% reduction in hospitalization of 1 day or longer for the unvaccinated. Vaccination was now expected to prevent hospitalization as well as for Delta, and previous infections reduced hospitalization risk by 60%: much better news.

In summary, everything that ICL knew about Omicron on 16 December was bad, but everything it knew on 22 December was good. Of course, by 22 December 2021, it was known that the government had not increased restrictions. It was soon going to become obvious that Omicron was milder, and that prior immunity, whether from vaccinations or prior infections, worked as well at preventing hospitalizations and severe disease as before. The function of ICL50 cannot have been to inform a past policy decision. It might have been to mitigate the negative impression created if ICL49 had been the team's last word on the matter. In any case, the ICL Covid-19 Response Team produced only one more report, its 51st being its last.

Taking the evidence from South Africa seriously necessarily implies some kind of descriptive exercise like modelling how the wave would play out in the UK if it took the same course as in Gauteng. If SAGE had taken the South African experience seriously, then it would have seen there was (as Streicher put it at the time) "no way that deaths in the UK will exceed 600 per day," let alone reach 6000 (Streicher 2021). There was just no viable explanation for such a radical difference between the two nations, even taking into account their many differences. Had SAGE not been so set on its methods, it might have avoided an error that probably contributed to the group being "disbanded"[6] and might have led to the great harm of an unnecessary lockdown.

3.10 Escaping Evaluation

When Omicron emerged, the example of South Africa made it possible to develop a descriptive model about the consequences of persisting with light restrictions in the UK. However, the predictive task facing SAGE in July 2021 concerned the result of a *change*: the lifting of lockdown restrictions. Descriptive models have limited, if any, ability to model the results of changes, since the curve just tells us what will happen if things carry on as they are. Descriptive approaches tell us nothing about why the curve is like it is, and thus cannot predict what will happen when the underlying situation changes. That is what mechanistic models are for.

However, there had been prior opportunities to assess the reliability of the methods used to generate predictions. Forecasts from SAGE's models, constructed retrospectively, could have tested whether these models made good forecasts of disease outcomes when real changes in exogenous factors (such as population movement) were used as inputs. Descriptive models can also find a use in this process of retrospective forecasting: to justify their continued use, mechanistic models might be expected to at least match, if not outperform, simple descriptive models. As one of the members of the modelling team at the London School of Hygiene and Tropical Medicine put it later (after the second error detailed here became apparent):

> It's worth thinking about how this process must have looked to an enthusiastic outsider: modellers have continued to deploy their essentially unverified models again and again at each new wave and reopening stage, seemingly oblivious to the accuracy of previous predictions.
>
> *(Hellewell 2021)*

Trying to determine the predictive performance of a model using retroactive forecasts is important, since the accuracy of counterfactual projections also determines the accuracy of the estimated effect of interventions. The most prominent evaluation of the effectiveness of the first UK lockdown was conducted by the same team that made the predictions about what would happen if there was no lockdown, which led to those lockdowns being implemented (Flaxman et al. 2020). Notwithstanding the unfortunate lack of independence, it would nonetheless have been easy to run a descriptive model on the available data at that time, to verify the assumptions used in the mechanistic model used to create projections for a no lockdown scenario. This was not done, and nor was it done a year later when the same team produced a fresh assessment (Mishra et al. 2021). This produced a much-reduced assessment of the lockdown effect. The reduction was attributed to anticipatory behaviour change. The paper is little cited, but its findings and explanation might have excited considerable scrutiny had they appeared a year before.

Nor, to our knowledge, has this simple exercise been done any time since, taking advantage of hindsight. We therefore decided to do it ourselves (Streicher, Broadbent, and Hellewell 2025). We retrospectively applied a current descriptive model for estimating R_t, EpiNow2, representing our best

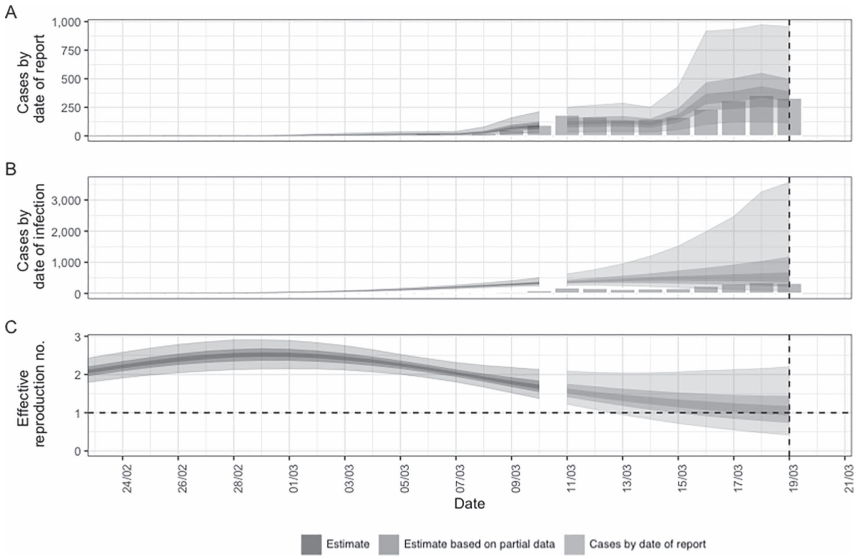


FIGURE 3.2 Current model for estimating Rt fitted to London case data as of 20 March 2020 (Figure reproduced from Streicher, Broadbent, and Hellewell 2025, 5)

current knowledge, to the situation as it was shortly before the first lockdown in London, as on 20 March 2020 (Abbott et al. 2024). The model suggests that R_t was well on its way to drop below 1 in London before the lockdown was introduced on 23 March 2020 (see Figure 3.2). This evaluative exercise was apparently not performed, and still had not been performed by SAGE in December 2021. Relying on a model without evaluating its past performance is obviously a poor practice of a basic kind.

3.11 South Africa's Approach

It is easy to blame poor predictive performance on the intrinsic difficulty of prediction. This is harder if others did better. The South African Covid-19 Modelling Consortium (SACMC) initially used mechanistic models to produce short-term forecasts that were inaccurate (SACMC 2020). They improved their performance over the course of the Pandemic. They did this by developing much stronger descriptive modelling tools and then deploying these alongside their mechanistic models. The process was obviously driven by a willingness to evaluate past performance and acknowledge error. In short, the SACMC exemplified both the aspects of methodological pluralism that we have just criticized SAGE for failing on.

South Africa locked down hard and early, with stringent restrictions in place from midnight 26 March to 31 May 2020. The SACMC released projections on 6 May but adjusted them on 12 June, roughly doubling the mortality for the Western Cape (SACMC 2020). At the time, they were relying on mechanistic modelling. Streicher contemporaneously applied a descriptive model based on fitting a log-ratio Gompertz curve to the actual mortality data.[7] As Figure 3.3 shows, this descriptive modelling exercise produced a much more accurate prediction. SACMC's original (pre-adjustment) prediction was also better than its adjustment, but still not very accurate as regards the timing of the peak.[8]

However, the SACMC learned from its first-wave experience. During the second wave, it eschewed mechanistic modelling and "made the difficult decision not to produce model-based projections," instead developing a "set of metrics that could detect and monitor the second wave" (Meyer-Rath et al. 2023). In our terminology, this means that the SACMC was not publishing the results of any mechanistic modelling (although no doubt there was modelling of this kind going on), but was working to better understand the statistical relationships between the different variables. Before we trust a mechanistic model about alternative scenarios, we ought to be confident of its ability to reproduce what

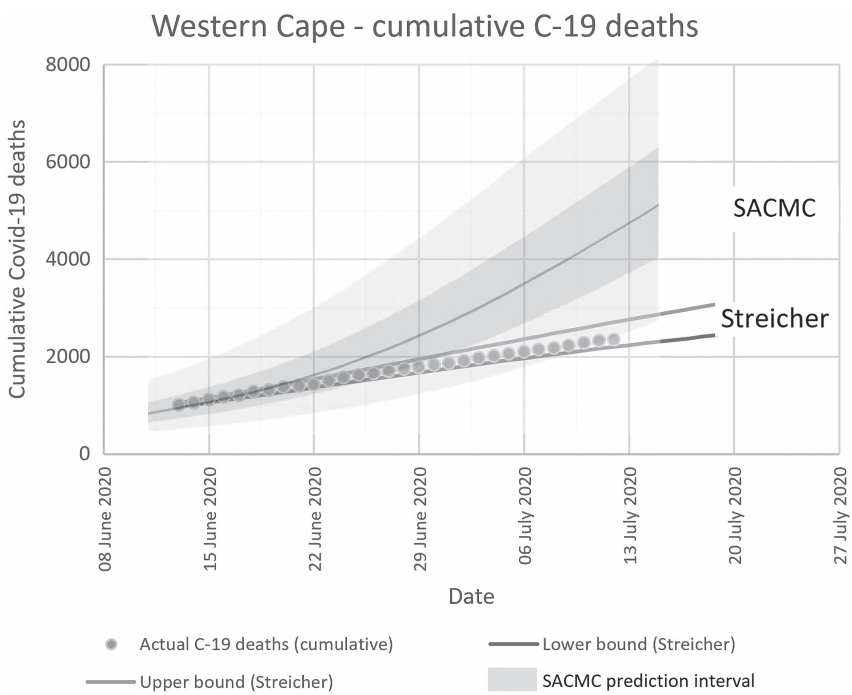


FIGURE 3.3 SACMC first-wave predictions for Western Cape, South Africa

actually happens, and this is supported by developing a detailed understanding of what actually happens.

The strategy paid off. By the third (Delta) and fourth (Omicron) waves, the SACMC was more confident, due to experience gained from monitoring the second wave, and produced both longer-term forecasts and short-term forecasts based on a combination of mechanistic and descriptive models. These performed well: "Finally, real-time analyses of the most important characteristics of the Omicron variant first identified in South Africa in November 2021 allowed us to advise policymakers early in the fourth wave that a relatively lower admission rate was likely" (Meyer-Rath et al. 2023, 2).

It is hard to oversell this success. In the case of Omicron, SAGE possessed not only more resources than its South African counterpart but also *more data*. Omicron had already peaked in southern Africa. The SACMC was, in retrospect, quite bold in advising its government not to implement further measures. SAGE had the advantage of seeing how Omicron played out in another country, *and* all the resources of one of the most powerful modelling communities in the world. But it set the data aside, squandering that advantage, and then performed far worse in what turned out to be a parallel prediction, that of forecasting what would happen in the country under unchanged restrictions.

SAGE never devoted its primary efforts to descriptive modelling, nor published the results of multiple approaches to modelling in the same report. (It did publish the results of multiple *models*, but of the same kind; and it handled discrepancies by introducing very large uncertainty intervals, as we have already seen, rather than follow the SACMC's example of admitting uncertainty and abstaining from making predictions.) The SACMC's cross-referencing of different approaches contrasts favourably with SAGE's unchecked confidence in a single approach. And SACMC's performance was better.

It was better because it was not methodologically rigid. In a paper we have already cited, the SACMC

> describes the 2-year process of continuously updating the models while collaborating with the diverse partners. . . . The aim is to highlight the dynamic, multidisciplinary nature of policy-driven modelling of an emergency in a country with severely constrained resources.
>
> *(Meyer-Rath et al. 2023, 2)*

There is a direct causal connection between the detailed development of a different methodology and the improvement of predictive performance. This is because that methodology offered a different strength that the mechanistic models lacked: it enabled a detailed understanding of what was actually happening, which is useful for evaluating, calibrating, and sense-checking mechanistic models, as well as for short-term forecasts into portions of the future – or indeed the present – which do not differ sharply from the context of the data to which the descriptive model has been fitted.

SAGE performed badly, on the other hand, because it was methodologically rigid. It published the results of multiple models of the same variety. It made no serious effort to reconcile their wide variance, and did not see this variance as reflecting any problem with the models, but as reflecting the inherent difficulty of being certain due to the operation of factors whose effects were unknown. In other words, it did not scrutinize its methodological approach, even when instances of that approach seemed to be producing very divergent results. It did not sense-check or seek to corroborate its predictions against simple descriptive predictions when making forecasts. It also did not retrospectively assess its mechanistic models against descriptive models, and outperforming such models would be a reasonable benchmark for the adoption of the latter as a basis for major policy decisions.

3.12 Sweden's Approach

Folkhälsomyndigheten (FoHM) is the Public Health Agency of Sweden. Sweden's approach, in a nutshell, was to estimate a Wuhan-based worst-case scenario for each region of Sweden. Wuhan was the region in China with the worst outcomes. Data from Wuhan were not very complete, but some things were known with reasonable reliability. One was the *clinical attack rate* (CAR), which is (laboratory-confirmed) cases as a proportion of the total population. There is an underlying assumption that those who are getting tested are mostly symptomatic: in other words, tests are only or primarily being administered to people arriving at clinics or hospitals with symptoms. This leaves several things undetermined: how many people are asymptomatic, how many are mild, how many experience significant symptoms but do not report them, etc. However, one can still apply the CAR from Wuhan to another population.

This exercise yielded reasonably accurate results for the crucial metric: peak ICU bed occupancy. The worst-case scenario was a peak of 12.5 beds per 100,000 population, which was within the capacity the Swedish healthcare system could prepare for (Berger et al. 2022). Figure 3.4 shows the actual ICU bed occupancy for different regions of Sweden against the FoHM's relatively simple model. The second highest regional peak was 9.1 beds per 100,000 in Stockholm, and the highest was 11.1 beds per 100,000 in Sodermanlands Ian. The model had provided a reasonable worst-case estimate. CAR was estimated at around 1%, and ended up at 0.8% after the first wave.

This does not prove that the decision not to lock down was correct, but it definitely does prove that the intense pressure Sweden came under to implement stricter policies was unjustified. That pressure was primarily based on Sweden's predictions being wrong. It came from the prospect of a devastating

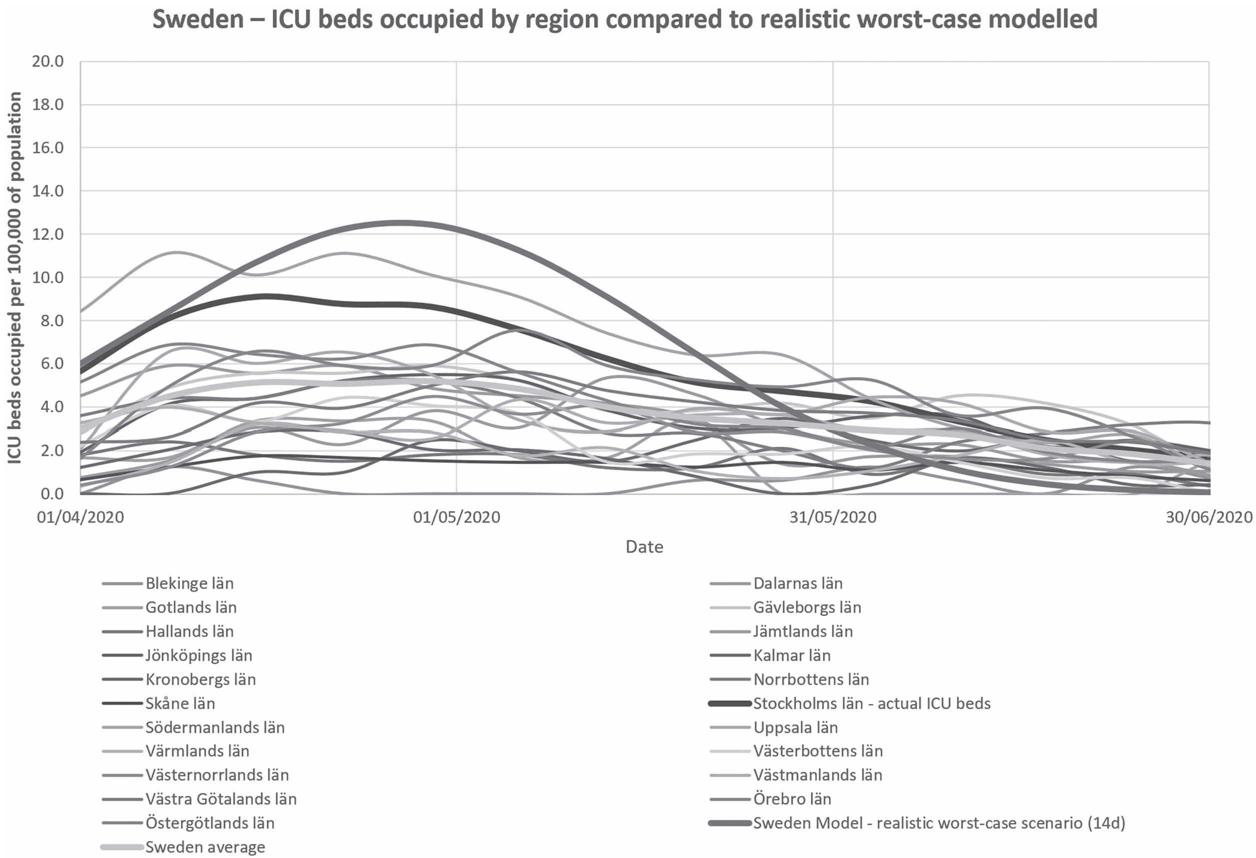


FIGURE 3.4 Sweden ICU bed occupancy against FoHM projections

epidemic in the country, one far worse than that predicted by Sweden's official modellers.

Sweden did not use the Wuhan-based worst-case scenario for want of modelling resources or expertise. The Nordic countries have excellent epidemiological expertise. They are a good place to do epidemiology because of the very large quantity of health data held centrally. They have the wealth to support large, research-intensive universities and a highly skilled epidemiological community. This includes epidemiologists with expertise in microsimulation modelling, of the kind upon which SAGE based itself in the UK, but which, on the other hand, was lacking in most countries globally. In an excellent paper, Till Grüne-Yanoff mixes philosophy with interviews to drive this point home (Grüne-Yanoff 2021). He quotes Lisa Brouwers, the head of the infectious disease-modelling unit at FoHM:

> there was still so much we did not know about how the virus was spread – how infectious it was. Simply taking the data from Wuhan and implement it in models for Sweden would render enormous outbreaks that were not realistic.
> *(Lisa Brouwers, quote in Grüne-Yanoff 2021, 45)*

The reason these enormous outbreaks were not realistic is that they derive from a feature of compartment models that is known to limit their usefulness: that they assume homogeneity. We will return to this in Chapter 4 in connection with the famous "R-number," which is an average that can conceal many different underlying realities (Yorke, Hethcote, and Nold 1978; Britton, Ball, and Trapman 2020). In the present context, it was pointed out early in the Pandemic that under homogeneity assumptions, compartment models would produce herd immunity thresholds in the range of 60–80% – that is, 60–80% of the population would need to be infected for a wave to start subsiding under no restrictions. Building variation in susceptibility and exposure into these models resulted in ranges of 10–30%, which is more realistic (Gomes et al. 2022). This point is familiar in public health, and is what Brouwers is referring to by "enormous outbreaks that were not realistic," but it was seemingly ignored by influential modellers during the Pandemic.

3.13 Avoiding Methodological Rigidity

What lessons can be drawn from comparing these cases of methodological rigidity and its absence? It is possible to identify several antidotes to methodological rigidity.

3.13.1 Clarity of Purpose

The first is clarity of purpose.

Modellers use their discretion to determine the appropriate model for a given exercise. This depends on the purpose. It is essential that modellers are clear as to the purpose of the exercise, therefore. Eric Winsberg and Stephanie Harvard argue that there is a moral component to the duty to be clear about purpose because models are inherently value-laden and may be used in decision-making (Winsberg and Harvard 2022).

There will always be aspects of a model that depart from reality. That is inherent in the idea of a model. Which departures from reality are acceptable will depend on the purpose for which the model is constructed (which, as Winsberg and Harvard argue, implies a value judgement). For example, a map showing the routes of all the underground trains in London may not be to scale, and it may not preserve much information about the geographical location of the stations. The standard map tells you nothing about whether one station is to the north of another, for example, and in some cases directly contradicts reality as to direction. It shows King's Cross as north of Angel, when the reverse is true. It also contains little information about relative distances, and none about absolute distances. However, it is useful for identifying the right train route between two points (provided one assumes that the time taken between two stations is roughly equal – an assumption that is not always satisfied but generally serves, and does not cause disaster when it fails). However, if one seeks to use the Tube Map to navigate London, one will get very lost, unless one gets very lucky.

Similar remarks apply to the predictive modelling of infectious diseases. For example, SIR models routinely over-estimate attack rate because by default they do not take account of the fact that populations are not homogenous and individuals do not mingle randomly, but contact each other in clusters – a point that was already established in the 1970s (Yorke, Hethcote, and Nold 1978). This makes them less suitable for resource planning purposes. SIR models can be adapted to account for this, but it is still an underlying tendency of this kind of model, and avoiding it depends on the success of the attempt to counteract it. On the other hand, SIR models are useful for explanation, understanding, and deriving the kind of predictive information that supports novel interventions – as we saw in discussing the famous case in the prediction that malaria could be eliminated without eliminating mosquitoes (see Section 2.3).

Agent-based simulation models have many parameters and require a large quantity and variety of data. They tend to be sensitive to the value of these parameters, meaning that errors, estimates, and assumptions will tend to have large effects. This means that their usefulness for constantly updated short-term forecasting is limited. They cannot be updated on a daily basis. It also means that their usefulness is limited in low-data situations. They are useful for comparing different policy options, and for longer-term forecasting about the possible effect of future changes – the emergence of a new variant, for example. But for telling us what will happen a week from now, they are not the approach of choice.

Curve-fitting approaches tend not to take account of differences between the source data and the target population, which makes them less useful for longer-term predictions. However, often these differences will not matter in the short term, but will only result in major divergences over longer time periods. They are relatively easy to perform, and this makes them useful for short-term forecasts. They can also be updated frequently, meaning that the curve can be modified to fit the target population over time, as happened with the IHME's model. But they are usually less helpful for longer-term forecasts or for predicting what will happen when novel events occur, such as the emergence of a new variant, nor for comparing policy options.

Wendy Parker has asked, "What does it mean for a model to be adequate-for-purpose? What makes a model adequate-for-purpose? How does assessing a model's adequacy-for-purpose differ from assessing its representational accuracy?" and has provided detailed answers (Parker 2020). For the present, practical purposes, however, we do not need to go so deep. The point is simply that the modelling, in a pandemic, is clearly done for a purpose; that modellers and those depending on models face a choice between models; that the adequacy of a model depends in part on the purpose for which it is being used; and that clarity about purpose is therefore essential.

3.13.2 Acknowledging Limitations

It is important, obviously, to acknowledge the limitations of one's methods. Mariusz Maziarz and Martin Zach point out that the outputs of agent-based models used for Covid-19 policy decisions do not measure up well when assessed by standard approaches to evidence because even after accounting for a long list of factors, "the question if all the key difference-making factors have been included in the model remains to be addressed" (Maziarz and Zach 2021, 3). Their point is not that there *are* limitations, but that they should be *acknowledged*, and that such models should be treated as just "one piece of the puzzle."

This is related to clarifying purpose, since "limitation" is a purposive notion: a limitation will always be relative to a given goal, end, use, etc. Thus, an effort to identify limitations also encourages one to acknowledge that some methods are better for some purposes than others. But acknowledging limitations pushes one further in the direction of methodological pluralism, since the most appropriate method for the *purpose* may still be seriously limited by the *circumstances*. As Maziarz and Zach point out, the limitations of a given method may differ in different circumstances:

> considering that epidemiological [agent based models] account for not only biological determinants such as infectivity but also social interactions that

> differ across the globe, the quality of evidence from [agent based models] must be assessed on a case-by-case basis.
>
> *(Maziarz and Zach 2021, 4)*

Acknowledging that your preferred methodological approach has limitations, and that these may be serious in some contexts even if the approach works well in others, will affect the confidence one has in one's predictions and may lead one to decide not to use a model that one would ideally like to use, as Sweden did. Ignoring limitations may lead to overconfidence or error, as it did in ICL9 and in SAGE's predictions of July and December 2021.

3.13.3 Triangulation and Sense-Checking

The epistemic merits of triangulation – the use of multiple independent models or methods to support robust inference – have been extensively discussed by philosophers of science, particularly in relation to climate modelling and simulation-based sciences. Philosophers such as Catherine Elgin, Jay Odenbaugh, Wendy Parker, Paul Teller, Michael Weisberg, and Eric Winsberg have explored in detail whether and how convergence across diverse models confers epistemic strength, whether disagreement between models signals underlying uncertainties, and under what conditions triangulation genuinely increases our confidence in model predictions (Elgin 1999; Odenbaugh 2011; Parker 2006, 2010, 2011; Teller 2001; Weisberg 2013; Winsberg 2010; Winsberg and Harvard 2024). There is a general consensus that triangulation is epistemically significant, even if philosophical accounts of how, why, and to what extent differ.

In the context of epidemiological modelling, the ready availability of different methodological approaches makes triangulation viable, as we have in effect demonstrated retrospectively. Triangulation can also be done prospectively, by using alternative models to produce contemporaneous forecasts. Where the models agree with each other, we have *triangulation:* different approaches yielding the same result. This is not proof of correctness, but it is strong evidence. Suppose that two different methods are used – such as extrapolating a curve from currently available data, and projecting it by modelling underlying interactions between individuals – and that they yield a similar result. If both are wrong, then why do they coincide? It is natural to think that this coincidence calls for explanation, and plausible at least to increase your confidence in the accuracy of both of them.

Conversely, a failure of triangulation is an important signal. Even if the descriptive approach is less informative about alternative scenarios, it can still be used to triangulate the output of a mechanistic model. – for instance, by indicating what will happen if we simply extrapolate the current situation. If the two models differ substantially, then we know that one of them, at least, must be wrong, and we must explain what we think has gone wrong before we can rely on the output of either model.

"Sense-checking" is a more pragmatic notion than triangulation, one that philosophers have not discussed. However, it is a widely acknowledged best practice among modellers and forecasters in various practically oriented fields such as epidemiology, climate science, and economic forecasting. Sense-checking typically involves comparing predictions from complex models against simpler descriptive or baseline models, empirical data, or common-sense extrapolations. Philosophers may not have given it a theoretical treatment, but it is extremely important in practice.

Sense-checking overlaps with triangulation. Within epidemiological modelling, sense-checking involves verifying complex mechanistic model predictions with simpler, data-driven descriptive forecasts as a guardrail against model-driven errors. However, it also goes beyond triangulation because it can be done retrospectively or in an ongoing way. As data become available, one can ask: do the predictions of yesterday's model about tomorrow still make sense, given what we have found out today?

For example, during the Pandemic, mechanistic models were commonly used to estimate critical care bed demand, deaths, and several other co-variables, under various scenarios. The relationships between these variables were estimated based on many factors, including how long people are likely to stay in a hospital bed or in a critical care bed, and what percentage of those in critical care would die. At the same time, data became available as to the *actual* relationships between these variables. It transpired that the crude ratios between certain key variables sometimes differed significantly from the ratios used by the mechanistic models. These major discrepancies were evident as early as ICL9. Disparities of this nature continued through the December 2021 SAGE models for Omicron. Observing the ratios between variables in the data and comparing these ratios to those in the models provides an easy way to sense-check the model as new data become available, one that does not require another modelling exercise, and does not extend to full triangulation. It is a simple check.

3.13.4 Evaluation

Where a mechanistic model is used to inform a policy decision, it is possible to evaluate its predictions in several ways. Prospectively, performing a descriptive extrapolation of the current situation will provide a kind of evaluation of the mechanistic model's projections under the current conditions, as we have already discussed in the context of sense-checking. This does not necessarily tell us which model is right, but enables some degree of evaluation.

Retrospectively, it is possible to check whether a model's predictions about the actual policy option that was selected were borne out by reality. We have shown how the decision *not* to make a policy change showed up SAGE's poor predictions in December 2021.

The evaluation of models was complicated by making model projections conditional on policy choices as opposed to contact rates. What models purported to tell policy-makers was what effect various polices would have on Covid-19 outcomes, but what they in fact modelled was the effect of various adjustments in social contact rates on Covid-19 outcomes. Whether a given policy would have the intended effect on social contact rates was entirely unmodelled – in fact, it was often an extremely crude guess. Eric Winsberg and Stephanie Harvard point out that this gives rise to enormous uncertainty about the effect of specific policies on pandemic outcomes because the effects of policies on the very same variables that are meant to represent those policies in the model may vary enormously from setting to setting (Winsberg and Harvard 2022). This also gave space for poor model performance to be explained away by appealing to the possibility that actual contact rates differed significantly from what was assumed for a specific policy. For instance, when rates fell prior to a lockdown, some argued that this was because of a "de facto lockdown," meaning a reduction in social contact rates in the absence of a lockdown, as will be discussed in Section 5.2.

It is also possible to seek data from places where different policies were enacted or different contact rates were achieved, in order to evaluate performance in counterfactual scenarios. For example, first-wave attack rates in Sweden ended at only 6%, varying from 15% in the Stockholm metro to well below 6% in rural areas. Sweden did not have stay-at-home orders or school closures for those under 15 years. Information from Sweden could have been used to evaluate and calibrate counterfactual models in other countries that matched the Swedish scenario, but where more stringent measures were implemented. Counterfactual scenarios were presented based on lower stringency levels, but these models were commonly not reconciled with countries such as Sweden, where such scenarios actually played out.

Further, after the first wave, it became possible to identify additional factors, not included in the initial models, but which significantly impacted the outcome. These included lived density, number per household, population health (obesity), poverty levels, and seasonal effects. Such an evaluation could have significantly improved the accuracy of disease modelling during the pandemic, but it was either rarely performed or rarely taken into account by the Pandemic Orthodoxy.

The evaluations that the Pandemic Orthodoxy employed used the same methods as the models they were evaluating, and we will discuss these in the context of lockdown in Chapter 5, where we will argue that these evaluations were distorted by the need to validate lockdown.

3.13.5 Presentation of Alternative Models

Another good practice is to present alternative models together. Again, the SACMC adopted this approach in South Africa, regularly reporting the results of different models. In the UK, however, SAGE did not do so. This is not because they were not receiving information from multiple models. There were modelling teams providing (on request) descriptive forecasts. This information went into the system, but there was often no trace of it in what eventually came out. As we have seen, SAGE's reports in July and December 2021 presented the results of four and two models, respectively, all of the same general kind, and did not indicate how their predictions were weighted – if at all – in coming to the eventual recommendations. This approach also prevents outsiders from performing their own evaluation and coming to a different conclusion. This may have been thought a good idea, since speaking with one voice in public may have been thought desirable. However, the subsequent Covid-19 Inquiry took a contrary view (Hallett 2024 – see in particular section 5), and so apparently did the South African team at the time.

3.13.6 Testing Assumptions

The first thing one does when assessing a model is ask what its assumptions are, and seek to evaluate these, especially with regard to the purpose of the model. This activity cross-cuts evaluation, sense-checking, and clarity of purpose. Assumptions ought to be clearly stated. This goes equally for all kinds of models – descriptive or mechanistic.

It is often impossible to know the values of all model parameters at the start of a pandemic. Modellers often assume values for variables. However, as we have already discussed, more information becomes available as time passes. The assumed values of all variables must be scrutinized as soon as information becomes available to update them. This provides a further basis on which to evaluate counterfactual models. Counterfactual scenarios from past models have a lasting effect on the believed effectiveness of measures imposed, and if the counterfactual scenarios are not evaluated in light of new evidence, then these persistent beliefs are unreliable.

One common problem arises when an argument appeals to a model in support, but the conclusion is baked into the assumptions of the model. For example, one might appeal to a model to support mandates rather than recommendations. But the model might assume – perhaps implicitly – that behaviour change only accompanies mandates. For example, ICL9 models the results of mandated behaviour changes (social distancing policies, etc.). Even though ICL9 acknowledges that significant voluntary behaviour change will occur, the model does not include this. That might be a reasonable simplification for some purposes, but it means ICL9 cannot be used to support regulations over mere recommendations.

Guaranteeing an outcome by baking it into the assumptions is a common risk in modelling. It is sometimes revealed when predictions turn out to be incorrect. As late as 2021 and 2022, poor modelling accuracy was explained away by pointing to people voluntarily changing their behaviour beyond what they were mandated to do (Whipple 2022). This is tantamount to admitting that the model did not support a regulatory intervention, as opposed to a recommendation, even though the same models were previously used to justify mandates.

Another example of an assumption that caused problems during the Pandemic was the homogeneity assumption in compartment and, to some extent, simulation models. We already mentioned this above as a limitation of SEIR models. Heterogeneity in susceptibility and exposure is known to significantly affect outcomes. If homogeneity has been assumed, or if efforts have been made to mitigate this default assumption, these assumptions need to be identified and evaluated as further evidence emerges.

3.14 Conclusion

Influential and well-resourced modelling communities made major predictive errors during the pandemic, some actual and some counterfactual. This was not due to lack of resources or data, but to methodological rigidity. There was (a) reliance on mechanistic approaches to modelling without due consideration of (b) the applicability of this methodological approach, (c) its vulnerabilities, or (d) its track record, or of (e) other relevant approaches. This methodological rigidity persisted for nearly two years and resulted in error.

The antidotes for methodological rigidity that we identify are as follows:

1 Clarity of purpose, because different methods are best suited to different tasks, among other reasons;
2 Acknowledging limitations, for obvious reasons;
3 Triangulation and sense-checking, for which different methods – especially simpler ones – can provide useful information;
4 Evaluation, for which, again, different methods can provide useful information, as well as emerging data;
5 Presentation of alternative models, which helps with several other items on this list;
6 Testing assumptions, which are present in any modelling exercise.

We specifically recommend a standard practice of *benchmark extrapolation:* producing a descriptive model extrapolating the current situation, as an accompaniment to any other modelling output. Perhaps this will not always be informative.

For example, while growth is exponential, it may not be possible to say on the basis of a predictive model when growth would stop being exponential. However, once the rate of growth starts to fall, it is possible to estimate the rate of fall in the growth rate (not in the epidemic wave bearing in mind that growth is still occurring) and extrapolate that fall. In other words, it becomes possible to estimate the timing and height of the peak, with increasing accuracy as time passes. This exercise will often provide the basis for many of the antidotes described above.

In this chapter, we have focused on methodological rigidity, arguing that it constrained scientific discourse and practice, and undermined pandemic responses. We now begin to move to the next element of scientific orthodoxy, namely, dogma. Methodological rigidity plays an important role in the creation of the first dogma we will consider, which concerns lockdown. The next chapter explains how epidemiological modelling created a strategic framework within which lockdown seemed unavoidable. Methodological rigidity plays a part in this explanation, and in the defence of dogma more generally, as we shall see.

Notes

1 This chapter draws upon material in (Streicher, Broadbent and Hellewell 2025).

2 For more on the distinction between contexts of testing and forecasting, see: (Broadbent 2013, 81–114).

3 So the sequence is as follows: 12 March Pueyo-Edmonds interview; 13 March SAGE meeting, where it was realized that UK case levels were significantly higher than previously thought; 16 March ICL9 published. This implies that the error regarding wave onset timing and peak timing was known at publication. It would have been better and reasonable simply to remove the months from the bottom of ICL9 Figure 2, since the point of that graph is to emphasise divergent outcomes, and this can be done without dates.

4 Robert Northcott also contrasts ICL's own use of descriptive and mechanistic models. He highlights the success of ICL's descriptive modelling of the new B.1.1.7 variant, compared to ICL9 (Northcott 2022). On this basis, Northcott argues for more contextual and informal methods, and for modelling that is empirically confirmed. These proposals align with ours.

5 Streicher's wording on social media was "the UK can expect 350 deaths per day at peak". However, in reality, he was presenting his worst-case scenario.

6 Graham Medley used the word "disbanded" in the UK Covid Inquiry. He was personally keen for SAGE SPI-M to continue.

7 The log-ratio Gompertz formula is $\text{In}\,(Y_n / Y_0) = ae^{-be^{-ct}}$, where Y_n is the cumulative number of COVID-19 deaths at time t, Y_0 is the initial number of COVID-19 deaths (at day 0), *a*, *b*, and *c* are curve-fitting constants, *e* is Euler's number (base of the natural logarithm), and *t* is time in days.

8 The Western Cape ended at 4,500 deaths after the first wave, compared to the SACMC's predicted 4,300. The model was, however, not accurate in terms of peak mortality rates (the steepness of the cumulative curve) or timing of the peak. The Western Cape peaked several weeks earlier compared to the 6 May 2020 model.

4

THE SCIENTIFIC CONSTRUCTION OF LOCKDOWN

4.1 Introduction

In the next chapter, we will make out the case that lockdowns were the subject of a scientific dogma. But lockdowns did not just happen. These were the truly unprecedented features of the Pandemic, which is why they call out for explanation. There had been pandemics before Covid-19, but the historical event of global lockdown was novel. What caused it? It is this question that we seek to answer in the present chapter.[1] Doing so shows how methodological rigidity, discussed in the previous chapter, led to the establishment of the dogma discussed in the next. It also enables us to explain why lockdown happened at all, which is important, since the content of a dogma needs explaining as much as the existence of the dogma. In our view, a scientific orthodoxy existed, which first created a conceptual framework in which lockdown seemed the only viable option, and then defended this policy against evidence that it was either less effective or less necessary than had been supposed. The present chapter concerns the first of these steps, and the subsequent chapter will address the second.

The origins of lockdown are to be found in a novel conceptual distinction between "suppression" and "mitigation" introduced in ICL9. This distinction was to become the linchpin of subsequent policy responses. Achieving suppression came to be seen as the only way to avoid unthinkable mortality. ICL9 does not use the word "lockdown" at all, but it introduces a threshold logic on which the only acceptable response was to go all out, since the difference between suppressing and failing to suppress was presented as radically divergent. Unfortunately, the suppression/mitigation distinction was conceptually fictitious, as we

DOI: 10.4324/9781003732495-5

show in this chapter. This was borne out empirically by the fact that outcomes in Covid-19 fell along a continuum, which we also show here. The strategic consequences were unfortunate: there was no Plan B for countries that found themselves locking down without achieving "suppression" as defined, and there was a resistance to considering alternatives that might have offered much of the benefit of lockdown for a fraction of the cost. This begins to encroach on questions of evaluation, to which we will turn in the next chapter.

4.2 The Suppression/Mitigation Distinction

The rationale for locking down arises (as we will show) from a distinction between "suppression" and "mitigation" that was afforded central strategic signficance by ICL9:

> Two fundamental strategies are possible: (a) mitigation, which focuses on slowing but not necessarily stopping epidemic spread – reducing peak healthcare demand while protecting those most at risk of severe disease from infection, and (b) suppression, which aims to reverse epidemic growth, reducing case numbers to low levels and maintaining that situation indefinitely.
>
> *(Ferguson et al. 2020, 1)*

The term "mitigation" was already in use, but the term "suppression" does not appear in SAGE minutes prior to the availability of ICL9. It also does not appear in the UK, CDC, or WHO pandemic planning documents (DH Pandemic Influenza Preparedness Team 2011, Qualls 2017, WHO 2019).

However, in ICL9, suppression became one side of a distinction which was presented as fundamental, almost as a conceptual truth: the suppression/mitigation distinction. That distinction was articulated in terms of the *reproductive number, R,* which, for a while, enjoyed a fame unparalleled by any other epidemiological metric.

At a given time t, R_t is the average number of people that each infected person infects. When R_t is greater than 1, each infected person infects more than one other person, meaning that the epidemic wave is still going upwards. When R_t falls below 1, the epidemic is in decline, and the wave has peaked – at least, has reached *a* peak, since R_t could rise above 1 again if something changes.

ICL9 makes the case for a strategy: that of taking whatever measures are necessary to bring R_t below 1, and keep it there. We call this the *suppression threshold strategy* (Streicher and Broadbent 2022). ICL9 does not make the case for any particular package of measures, and certainly not lockdown, which it mentions only briefly and does not model. ICL9 only models NPIs *short of* lockdown. Nevertheless, ICL9 laid the basis for lockdown, as we will now see.

4.3 Threshold Logic

Typical policy-making involves evaluating the marginal benefit and cost of adding or removing individual components of a package of interventions. This implies some form of benefit–cost analysis. An individual intervention will only be acceptable if the benefit of implementing the intervention exceeds the cost, and its inclusion in a package will likewise depend on the marginal benefit and cost of including the intervention.

This approach assumes that benefits (and costs) lie on a continuum. However, sometimes there may be a threshold, such that reaching that threshold confers enormous benefits relative to missing it. Then a different logic might come into play. Hugely costly measures with tiny marginal benefits may be justified if they contribute to the threshold being reached, because in that event, the entire benefit of crossing the threshold is gained. In this situation, it may well be reasonable to forget about costs and make maximal efforts to attain the threshold, since the benefits of doing so are so huge – or the costs of failure are so dire.

The putative suppression threshold of $R_t = 1$ functioned in this way to justify lockdown. ICL9 presented the dire consequences of failure to suppress. While some consequences were worse than others, they were all unacceptably bad. A universe of policy options was thus condensed into a strategic binary: aim for suppression, or do not. The idea of weighing up different policy options, all of whose consequences are horrendous, is apt to seem like an irresponsible waste of time. The only acceptable strategy, as ICL9 puts it, is to suppress the virus.

This is how a report that does not advocate for lockdown laid the basis for lockdown. Unfortunately, the basis is invalid. The threshold between suppression and mitigation in terms of R_t is illusory, as we will now see.

4.4 There Is No Suppression Threshold

In the abstract, it may seem almost a matter of definition that measures designed to force R_t below 1 are seeking to *suppress*, while those seeking to reduce it but not necessarily below 1 are seeking merely to *mitigate* it. However, this is not true, let alone true by definition. From an early point in an epidemic wave, other things being equal, R_t falls inevitably. The question is merely how fast. All strategies seek to accelerate R_t's descent, and, as we shall now see, there is no "fundamental" strategic distinction between different rates of descent, still less between different times at which R_t falls below 1.

R_t is in decline from an early point in an epidemic wave and eventually falls below 1 in all finite populations, since infections cannot grow to infinity. Thus, a timeframe is needed in order to give sense to this way of drawing the distinction between suppression and mitigation. The distinction between suppression and mitigation, if defined in terms of reproductive number, is a *temporal* distinction.

However, this creates considerable difficulties for the idea that it is a fundamental strategic distinction. That is because it requires deciding on a timeframe within which R_t must fall below 1 after a measure is implemented, if we are to count it as a successful suppression. This, however, is very hard. We do not expect an intervention to have a measurable outcome in case count on the very day it is introduced, because there is an incubation period, about which there may be considerable uncertainty. There are further lags until case ascertainment, hospitalization, and death, all of which may also be uncertain. There is also likely to be some fuzziness around the exact start point because people may begin adjusting behaviour before a lockdown (or, in not a few cases, engage in an intensified programme of social engagements). Moreover, these may not be sources of actual variation in the underlying reality, as well as epistemic uncertainties: people may change their behaviour at different times, the disease may progress at different rates in different people, etc. In short, the expectation timeframe for evaluating the effectiveness of an intervention on any of the outcomes of interest, including R_t, is a complex matter of judgement under uncertainty. Far from being a matter of principle, it seems to be a thoroughly practical judgement. This undermines the suggestion that it can form the basis of a "fundamental" strategic distinction.

4.5 Paradoxical Consequences

The problem with specifying a timeframe goes beyond mere practical difficulties, however. The deep problem is that how quickly R_t falls below 1 is simply not a surrogate for any reasonable notion of "suppression," and if we treat it as such, we get bizarre and paradoxical consequences.

For instance, a short, sharp wave may result in R_t falling below 1 sooner than a long, flat wave, even if many more people die in the short, sharp wave. An intervention might cause a wave to be longer and flatter, and save many lives, yet postpone the dropping of R_t to below 1 to later. This would imply that, absent the intervention, the disease burned through the population extremely fast so that susceptible people were swiftly exhausted (i.e. immune or dead), while the intervention, thanks to its success, preserved many people much longer in a susceptible state. On this basis, doing nothing might even count as the suppression measure, depending on what timeframe one settles on for one's measure – in this case, inaction – to bring R_t below 1. Figure 4.1 shows how R_t can fall below 1 faster in a wave that is nonetheless obviously less desirable. (Bear in mind that this is merely a *conceptual* point, and does not show that this *will* happen in any given situation. We are merely showing that bringing R_t below 1 sooner rather than later does not *necessarily* mean reducing attack rates.) If one has selected sometime between the two peaks as one's criterion for suppression, then doing nothing will count as suppression, while intervening will count as mitigation, which is obviously ridiculous.

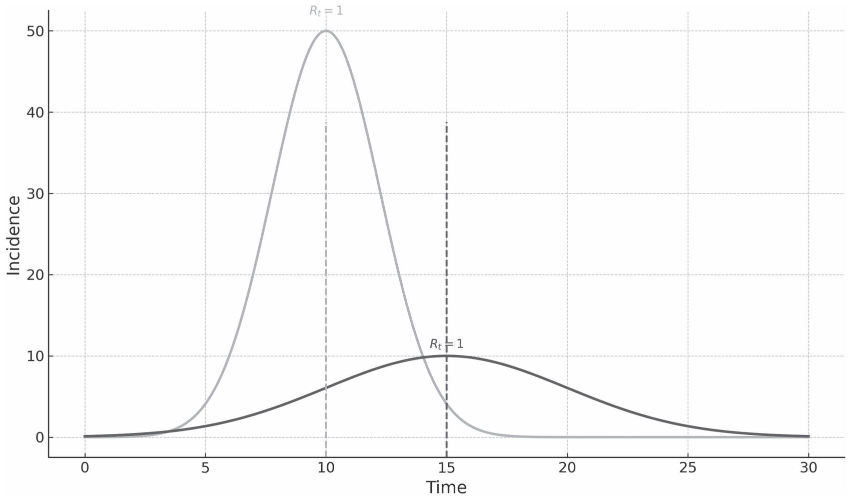


FIGURE 4.1 Early peak does not always indicate a positive outcome

For another example, the exact same package of measures (e.g. a "hard lockdown") could count as mitigation implemented early, because it might reduce R_t but not bring it below 1 in the specified timeframe. But the same package might count as suppression if implemented later when R_t was about to fall below 1 anyway (Wood and Wit 2021). Then the contribution made by the intervention package might be just enough to bring R_t below 1 within the timeframe (when otherwise it might have been a day later, say), and thus the interventions succeeded in suppression according to this way of viewing the matter. However, this is quite paradoxical. Not only are they the exact same package of measures, but in the scenario where they are introduced early (and thus count as mitigation by this definition), they save many more lives. Figure 4.2 shows this situation.

The same figure could be used to illustrate how a milder package of measures introduced earlier might save more lives than a lockdown coming too late – a situation which nearly everyone agrees was unfortunately common during the Pandemic. It also shows how a late lockdown can be wrongly credited with bringing about a peak simply because the peak follows hot on the heels of the lockdown.

The upshot is that if we focus on R_t out of context, then we reach the conclusion that mitigation can sometimes save many more lives than suppression, which is paradoxical. We must also be willing to accept that the very same package of measures can be either suppression or mitigation, depending on when it is introduced. These consequences contradict the idea that there is a fundamental strategic distinction in play.

Defining strategies in terms of the absolute value of R_t taken out of context does not help clarify the conceptual and strategic situation. The overriding measure of

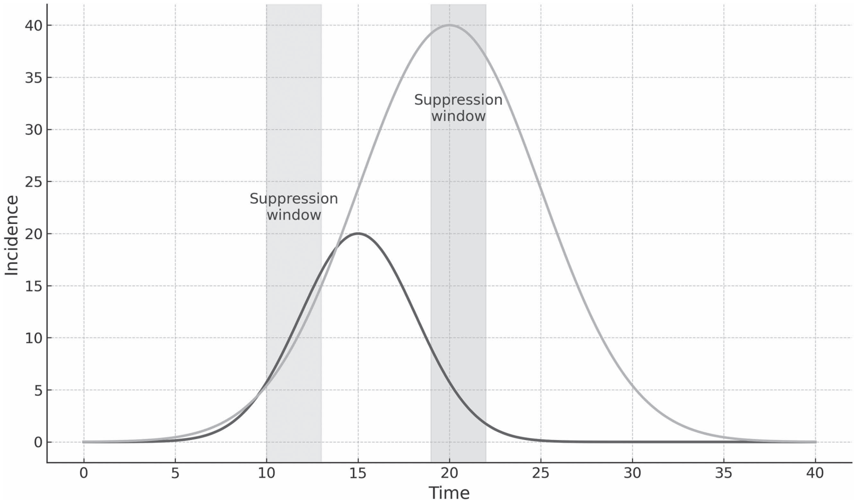


FIGURE 4.2 Early interventions are more effective yet less likely to count as "suppression" on the ICL9 definition

success is eventual attack rate (AR) as applied to cases, infections, or deaths. As we have discussed, it is possible to make predictions relying on attack rate without an estimate of R_t. It is not possible to run a compartment or simulation model without referencing R_t. The importance of R_t in mechanistic models might explain its centrality in ICL9, but it is not reflected in any real-world strategic distinction.

4.6 A Better Definition of "Suppression"

The term "suppression" is still useful, we believe, when defined as follows.

> Suppression is reducing population-level attack rates for a fixed period (e.g. until widespread vaccination).

On this definition, suppression is a matter of degree, not of crossing a certain threshold, and it is relative to a time interval, rather than either an indefinite state or a quantity at a point in time. On this definition (unlike on ICL9's), it makes no sense to say that a disease is suppressed at any given point in time. The question is to what *extent* ARs were reduced over a given period of time. In the Pandemic, the endpoint of this time period was widespread vaccination.

Because suppression in our sense is a matter of degree, countries can decide what costs are worth paying for what degree of suppression. A country could decide that enough suppression could be achieved without a lockdown, and that a lockdown represents an unwarranted further cost, even if it would achieve a higher level of suppression. We can also say that two countries have different levels of success in achieving suppression, depending on the level of suppression they achieve.

There is still a contrast with measures that do *not* seek to minimize population-level ARs. These might include, for example, the "Focused Protection" strategy associated with the Great Barrington Declaration (see Chapter 8). The goal of that strategy was to protect vulnerable people, not to control ARs in the population at large, and the unpredictability of vaccine development was part of the rationale. By contrast, Sweden followed a suppression strategy consistently, on our definition, since it implemented measures to reduce ARs, even though it never locked down.

This definition of suppression permits us to define a *suppression proportion*, which we will do in the next section to evaluate the relative success of different countries at achieving suppression.

4.7 Suppression Outcomes Fell on a Continuum

If the fundamental distinction between suppression and mitigation strategies in terms of seeking to force R_t below 1 is indeed a chimera, then one would expect the outcomes of stringent population-level control measures, including lockdowns, to fall on a continuum. To assess whether this was in fact the case, we considered the top 46 countries by *GDP* per capita (as they were more likely to have the means for accurate surveillance). The pandemic was divided into two equal periods: March 2020 until the end of June 2021 (16 months), during which all countries applied restrictions to minimize ARs prior to vaccination rollout; and July 2021 until 1 November 2022 (16 months), during which all countries started lifting restrictions, allowing the virus to spread. Although different countries vaccinated at different times, our concern here is to gauge the level of suppression (on our definition, i.e. the extent to which population ARs were minimized), while most countries in the group were still aiming at this. Hence, the cut-off period of 1 July 2021 was chosen as being just before the first countries started to relax all restrictions.

We then calculated a *suppression proportion* (SP) of the pandemic during this interval as follows, where C_{SUP} is the cumulative case count on 1 July 2021 (i.e. cases during the suppression period) and C_{TOT} is the cumulative cases on 1 November 2022 (i.e. total cases).

$$SP = \frac{(C_{TOT} - 2C_{SUP})}{C_{TOT}}$$

SP of 0% means that half of the total cases happened in the first 16-month period, and half of the cases happened in the second 16-month period. In that instance, there is no difference between the period in which suppression was being undertaken and the period in which it was not. The proportion of cases that have been suppressed is 0% if there are no other factors in play. SP of 100% means that 100% of cases happened in the second 16 months (i.e. outside of the suppression period), and thus the proportion of the total that was suppressed was 100%, if no other factors are in play.

Naturally, many other factors are in play, however. The timing of waves and the emergence of new variants, for example, could result in differences even in a hypothetical situation in which the exact same interventions remained in place across the entire 32 months. Nonetheless, our goal is not to draw a causal inference but to establish a *correlation* between the timing of cases and eventual mortality. We are seeking to establish whether there is a correlation between the extent to which cases were "postponed" until after vaccines were available and eventual mortality. Using SP for this purpose does not imply any causal assumptions.

In Figure 4.3, 100% on the *x*-axis indicates a maximal level of suppression as all C-19 cases happened during the second period, after 1 July 2021, when restrictions were relaxed. 0% on the *x*-axis indicates equal cases in the first period compared to the second period (i.e. there was little, if any, suppression achieved during the first period). Total C-19 deaths by population for the whole period

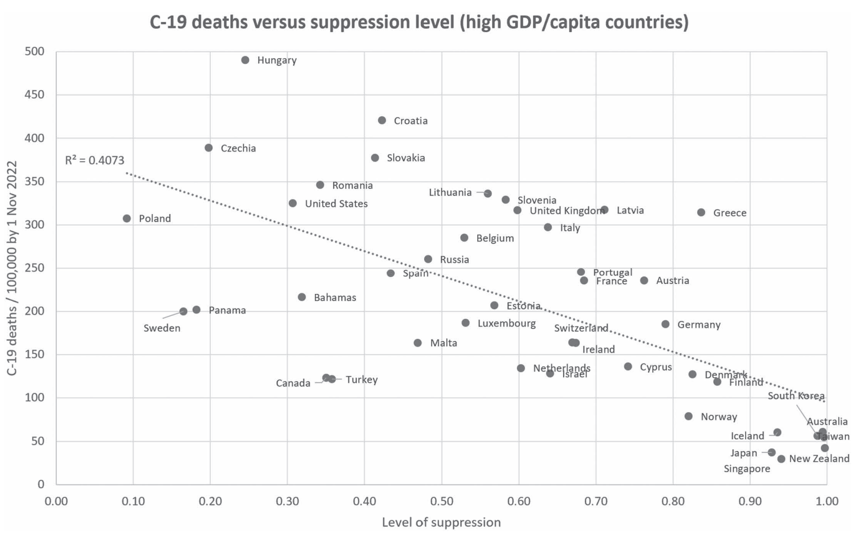


FIGURE 4.3 Covid-19 deaths and suppression levels in 46 high-income countries (reproduced from Streicher and Broadbent 2023)

(up to 1 November 2022) were plotted on the *y*-axis. The aim here was to see if there was any association between higher suppression levels and lower ultimate C-19 deaths, that is, are there signs that delaying infections had a benefit in terms of reduced C-19 fatalities. Case and death data by population were obtained from the 91-DIVOC data visualization tool using data from Johns Hopkins University (Fagen-Ulmschneider 2024).

Table 4.1 summarizes the results. There is a wide range of ultimate outcomes (200–491 C-19 deaths/100k) for those countries with low levels of suppression (0–30%), indicating that there must be other confounding factors that affected ultimate C-19 deaths. In contrast, all countries with very high levels of suppression (>90% for cases) ended with very low C-19 mortality levels (<61 C-19 deaths/100k). This category includes the countries that succeeded in containing the virus for long periods prior to vaccination, so that most of the population is completely virus free. (We define containment in Section 4.10.) This is to be expected since if the population is largely virus free, then the other factors resulting in varying outcomes cease to be relevant.

Overall, a reduction in ultimate Covid-19 deaths was associated with an increase in suppression levels (see Figure 4.3), albeit not strongly, presumably due to the existence of other relevant factors, as recently mentioned (correlation coefficient = 0.4073; it is plausible that controlling for strong confounders like poverty or vaccination level would increase the correlation). This was as expected and suggests that delaying infections until after vaccination had a beneficial effect in terms of ultimate Covid-19 deaths.

However, considering Figure 4.3, it is apparent that suppression is not a threshold phenomenon. The outcome (mortality) of the majority of high-*GDP* per capita countries lies on a continuum, depending on the level of suppression achieved. We only see consistently low C-19 deaths when suppression levels exceeded 90% for a handful of countries, mostly island nations. These nations achieved *containment*, which is different from suppression, and constitutes a true threshold (discussed later in this chapter in Section 4.10).

Our analysis confirms the widespread view that reducing ARs prior to vaccination improved eventual mortality. If suppression is defined this way, then

TABLE 4.1 Covid-19 deaths by suppression level.

Suppression level	*C-19 deaths/100k*
0–30%	200–491 C-19 deaths/100k
30–50%	130–421 C-19 deaths/100k
50–70%	130–340 C-19 deaths/100k
70–90%	80–320 C-19 deaths/100k
90–100%	29–61 C-19 deaths/100k

suppression works. However, there is no magic threshold of suppression that yields a step change in mortality. If suppression is defined this way, then it does not work. This confirms the outcome of the conceptual finding of the previous section that there is no suppression threshold. Outcomes lie on a continuum: the more suppression was achieved, the better the eventual outcomes. Thus, no basis remains either in theory or practice for thinking there is a fundamental strategic distinction between suppression and mitigation strategies.

4.8 The Problem of Diminishing Returns

We turn now to consider the consequences of the erroneous idea that there was a suppression threshold, and the misguided strategy we have called the Suppression Threshold Strategy or STS (Streicher and Broadbent 2023) – that of trying to attain it with little or no regard for cost.

One serious consequence of STS is that it has led countries to incur much higher costs than would otherwise have been accepted for given benefits. We say the costs are *much* higher (and not simply higher) because the marginal benefit of each subsequent measure tends to reduce, while the marginal cost tends to increase. As one economic analysis puts it:

> As policymakers tighten the noose on the economy, they attain progressively smaller healthcare benefits while generating progressively larger income losses.
>
> *(Camera and Gioffré 2021)*

This is compounded by the fact that the last measure to be introduced – a full lockdown – has the highest intrinsic cost of all, at any given stage of introduction (i.e. it would be the costliest first choice, the costliest second choice, etc.). That is exactly why it is introduced last.

Diminishing returns occur when any intervention reduces infection rates by some factor (rather than by an absolute amount, for example, if all inhabitants of care homes were protected from infection). Imagine population-wide measures A and B, both of which would, if introduced, halve the number of infections on a given future day. A package consisting of just measure A might then halve the number of infections on that day from, say, 100 to 50. Including measure B in this package would bring that number down to 25. Additionally, including measure B in this planned package would bring a marginal benefit of 25 prevented infections, half that of measure A, even though both are equally effective. The returns diminish for each additional measure we build into the package. This also means that if we decide on including B in our package, and then consider whether to include A, we achieve the reverse result.

These considerations apply even when one is considering inclusion of measures in a package that will introduce them simultaneously, but obviously it also applies even more so to the contemplation of adding additional measures to some that have been introduced at an earlier time, since the later a measure is introduced, the less its potential benefit will be (since the greater the number of infections that will have already occurred).

Costs, on the other hand, will not diminish with additional measures. The cost of closing schools is not mitigated by closing restaurants (either as part of the same package or as an additional subsequent measure). On the contrary, costs may tend to compound. For example, if a person is able neither to meet friends nor to send their children to school, then their mental health is likely to suffer from both the inability to access support networks at a time when they are particularly needed, and thus to have a greater effect than the sum of either stressor alone. Costs will not *always* compound – perhaps not having to deal with one's boss on a personal basis is some relief. However, the experience of many who experienced lockdowns was that the combined effect on finances, careers, business, personal relationships, children, and a host of other factors was considerable. The costs also compounded more lower down the social hierarchies, as we discuss in Chapter 10.

Diminishing returns are not at all unusual in policy construction. An optimal balance may be elusive, but the process of weighing up marginal benefits and costs to find an acceptable solution is a familiar one, even if it is complicated, and even if there are different ways to approach it.

However, if a strategy is premised on hitting a certain threshold, then diminishing returns are not relevant until the threshold is crossed. The entire benefit of surpassing the threshold is to be played for. Of course, beyond that threshold, the problem of diminishing returns may arise, but while one pursues it, the important question is whether a measure is necessary in some package that will be sufficient for crossing the threshold. A measure might be the straw that breaks the camel's back, so to speak, and it might make all the difference between success and failure despite having a negligible effect on infections. This could be true even if its cost is very high.

If there is no threshold, as for suppression, then the problem of diminishing returns is wrongly set aside. As discussed, the entire benefit of surpassing the threshold stands to be gained, and so even a small marginal benefit might be justified if it is the straw that breaks the camel's back. But if there is no threshold, then the crucial calculation of policy-making has been ignored. In the situation where there is no threshold, it is clear that the optimal strategy cannot be arrived at simply by ignoring costs, but will involve a judgement of the benefits and harms of each measure. It would be a remarkable piece of good fortune if the optimal policy were arrived at without this task being undertaken.

If there is a threshold, then diminishing returns do not present the same problem, but they become relevant in a different way: they affect the stakes of failure, to which we now turn.

4.9 The Lack of Plan B

What should be done in those many parts of the world where containment is unachievable, and suppression turns out not to be a true threshold (and perhaps would not be achievable even if it were)?

A serious strategic consequence of STS is its silence on what to do if a suppression effort fails to tip outcomes over a threshold. This situation may arise because the threshold turns out to be an illusion, and it becomes obvious that outcomes are falling on a continuum (as was the case for C-19 suppression). Or it may arise because the threshold, while real, is unattainable (as containment might be in many places).

We have already seen that the logic of justification when one is trying to attain a threshold differs from the more familiar process of weighing up benefits and costs. For example, imagine that a country has achieved near-containment, but schools are still open. Suppose that closing schools will result in a relatively small reduction in transmission. Nonetheless, it might be justified if closing schools enables the country to reach the containment threshold. Then schools can reopen, and much more besides. However, if school closures do not tip the balance in favour of a dramatically improved outcome, then their modest direct contribution to reducing transmission might not be sufficient to outweigh their huge cost in welfare (Munro and Faust 2020; Viner et al. 2020).

It is therefore important to be clear whether threshold logic or more familiar benefit–cost logic is in play. In particular, it is important to be clear when an attempt to reach a threshold has failed, and thus when the logic needs to change. The result may be a change in policy, potentially a dramatic one.

For example, if it becomes apparent that a country will not attain some threshold – be it the putative suppression threshold we have criticized, or the true threshold of containment – that will usually mean that restrictions should be relaxed – *even while cases are still rising.* Once it is clear that a threshold cannot be reached, it becomes necessary to assess the additional benefit against the additional human cost of including each measure in the overall intervention package. The result might be that the additional benefit of including a measure in an intervention package is rather modest, while its cost is rather high, and thus it is not worth including. That modest benefit would have been worthwhile if it had enabled us to cross a threshold and enjoy a very large benefit. But if the threshold has been missed or was an illusion, then the situation changes completely, and we are back to weighing up costs and benefits.

To continue with the school closure example, suppose schools are closed initially, but it later becomes clear that a threshold of containment is not achieved (or that suppression outcomes are lying on a continuum). Previously, school closures were justified despite their high cost because including them enabled the whole intervention package to cross a threshold that resulted in far lower mortality. However, this no longer being the case, the justification for incurring these costs no longer exists. We are obliged to consider whether the benefits of closing schools in terms of eventual mortality outweigh the harms of doing so. Since the latter are considerable while the former may be smaller relative to some other measures (hypothetically), the school closures may become unjustified. Then the appropriate response may be to reopen schools, *even if cases are still rising or start to rise subsequently.*

This may feel quite paradoxical. "Common sense" may suggest that minimizing transmission remains the appropriate goal even if we are not getting the results we hoped for. Imagine braking hard to prevent a collision. You will continue braking hard even when it is obvious that a collision is inevitable because that will reduce the impact.

However, common sense is a poor guide to public health strategy, and can yield conflicting results. Imagine you are running for a train, and you realize you will just fail to reach the doors in time. You could leap for the train as it passes, or keep running towards your destination as hard as you can. However, you would probably instead wait for another train. It must be acknowledged that epidemic suppression is not quite like this, because there is *some* benefit to retaining costly measures. Nonetheless, the step change that was hoped for can be missed (like a train), and it is in the nature of a threshold that missing it makes a big difference. This implies that you will need to reconsider, and it is likely you may need to change course.

Where threshold phenomena in epidemic response strategies are concerned, Plan B may be quite different from Plan A. Future pandemic strategies need to appreciate this point and plan for scenarios in which the threshold is missed, for whatever reason. In the C-19 pandemic, evaluation of the comparative benefit of different packages was initially discouraged because the consequences of failing to hit the (putative) suppression threshold were depicted as dire. The impression was created that trading off costs and benefits was a waste of valuable time and that all strategies short of maximal suppression would have disastrous outcomes. In effect, all "mitigation" strategies were bundled together along with doing nothing at all:

> our counterfactual of an unmitigated pandemic clearly demonstrates the extent to which rapid, decisive and collective action can prevent billions of infections and save millions of lives globally.
>
> *(Walker et al. 2020)*

Although they acknowledge the existence of "challenging decisions faced by all governments," the implication of this sentence is not one of trade-off, complexity, and nuance. Yet, it was quite foreseeable that many nations would never achieve anything like a 75% reduction in social distancing and that many suppression efforts would fail. The report just cited, Imperial College Covid-19 Response Team Report 12 (ICL12), purports to inform global policy, with a special emphasis on low- and low-middle-income countries, but in fact it fails to provide useful guidance, since it primarily extols the benefits of measures too "decisive" for such countries ever to implement. It does not offer a Plan B for countries that could not afford to bet everything on hitting a distant threshold. The lack of guidance for Plan B was extremely unfortunate and needs to be addressed in planning for future pandemics.

4.10 Containment

There is a real threshold for control measures to aim at. However, it is not suppression in terms of R_t or any population-level AR. It is *containment*, defined as follows. This implies preventing spread at an early stage, so that the vast majority of the population remains uninfected. When executed successfully, success is indicated by daily case counts at or close to zero in the presence of a comprehensive testing programme. This is a true threshold, since case counts either are or are not at this low level, and if they escape this low level, then they will quickly reach a point where that level is no longer attainable. In the limit case, containment everywhere will lead to complete elimination, as happened with smallpox. However, this requires simultaneous global success or complete and permanent geographical isolation.

Containment is very different from suppression in either of the senses we have discussed (Baker et al. 2023). A successful containment strategy will prevent the infection from ever getting much of a foothold. R_t may rise above 1, but the affected population will be tiny, and it will quickly fall. Any outbreaks result in action that is early enough to prevent them from spreading far at all. A successful containment strategy results in long periods of time during which the disease is absent. This permits light restrictions during those periods. A lockdown can form part of a successful containment strategy, but if the strategy is really successful, such that it is implemented when outbreaks are small, the lockdown ought not need to be in place for long. Very tight border controls are also important if the disease is present elsewhere in the world.

Containment forms a true threshold that, if attainable, makes a threshold-based justification available for all-inclusive efforts that include interventions with less attractive individual marginal benefits and costs. Unfortunately, during Covid-19, there were very few countries that stood a chance of approaching containment. This was well understood, and few countries aimed at it, thinking there was a more attainable threshold of suppression that was within reach. Moreover, containment is a risky strategy because the threshold is easily missed. Then, one has implemented very costly measures without the hoped-for benefits, and in a multi-wave scenario, the results of widespread infection will be bad.

4.11 Containment Is Difficult

Containment is, as we have said, a true threshold. You either succeed in containing and eliminating the virus, or you do not. However, strategies seeking to achieve containment suffer from two major problems, and these must be considered before lockdowns are adopted in order to achieve containment. These problems are *non-universality* and *instability*.

Some diseases are better candidates for containment than others, and Covid-19 is one of the others. Eliminating Covid-19 is very difficult. Not a single country in Europe, Africa, North America, or South America succeeded in containment at any point during the pandemic. Only a small number of countries, all in East Asia or Oceania, succeeded in containment during any part of the pandemic.

Containment of Covid-19 seems not to be a realistic choice for all countries. The great difficulties that countries in Europe and the Americas experienced with even achieving target levels of suppression show this. The countries that managed it share interesting features. Oceania consists of island nations, which facilitates border control measures, and are typically sparsely populated, and some with high income, which makes compliance more feasible. These countries also tend to have high obesity levels and high average age, and, in the case of New Zealand, low ICU bed capacity, all of which meant that containment was a highly desirable objective. New Zealand provides the clearest example of successful containment, spending 35 days under lockdown (Alert Level 4) from 23 March 2020 until 27 April 2020, and the capital city, Auckland, spending another 35 days under lockdown (Alert Level 4) from 17 August 2021 until 21 September 2021.

However, even for Oceanic countries that could achieve containment, a second issue arose: instability. With each repeated wave of infections, containment must be successful. If it is not, then much of the previous effort will have been in vain. The ensuing wave is often bigger than the first waves in other, non-containment contexts (Findl and Suárez 2021). There is an endpoint, of course: vaccination. Risk is often conceptualized as the product of the probability of a negative outcome and its gravity. Until vaccinations are available, both the probability and cost of containment failure – in short, the risk – grow with each successive wave.

At the extremes, lockdowns either result in quick containment, or it is obvious from the start that the containment is impossible, and the lockdown ought to be lifted early. One problem for containment strategies arises where containment is almost attainable, and appears to be, leading to a lockdown that is either futile or too long. Another serious problem arises where attainment is achieved in one or more waves, but not in a later wave. In this scenario, a country or a city will include a very large number of people who are highly vulnerable because they have no prior exposure to the virus. The area may become “locked in a lockdown,” facing significant growth in infections if the lockdown is lifted, yet failing to reap the benefits of containment from the lockdown.

The city of Melbourne in Australia provides the clearest case study of a scenario where containment becomes increasingly difficult and increasingly short-lived, ultimately leading to failure, which neutralized many of the benefits of the early success. Melbourne experienced 263 days in 565 under hard lockdown between 30 March 2020 and 21 October 2021. That is nearly half the days in that period. The city successfully contained the virus five times, in five separate lockdowns. The second lockdown, however, was applied late and lasted 111 consecutive days (from 2 August to 9 November 2020). The third, fourth, and fifth lockdowns were all implemented much earlier, before daily cases exceeded five. These lockdowns all succeeded in containment in less than 14 days. However, after the fifth lockdown, Melbourne only lasted nine days without any cases. On 5 August 2021, Melbourne entered its sixth lockdown, and this lockdown never succeeded in reversing epidemic growth. After 78 days, Melbourne lifted the sixth lockdown while cases were still rising. The reasons included both discontent of citizens and some degree of vaccine uptake, but the situation was far from ideal because vaccine uptake was slow. Whether a lockdown would have been continued (as it was in the second wave) if vaccines were not available at all is a nice question.

4.12 Containment Is Worthwhile Only Where It Is Difficult

The instability of containment appears not to have been universal. East Asian countries were in a position to achieve containment because of their early response and the effect it appeared to have. Moreover, they seemed able to do so without lockdowns. South Korea and Japan both achieved containment in the first wave without a national lockdown. However, they drew a different conclusion from the Oceanic countries, deciding that in future waves it would be acceptable to have a low level of cases for the sake of avoiding strenuous efforts to contain. This was because the relative ease of attaining containment meant that it was not unstable in that context.

The reasons are not transparent, but may have something to do with low obesity. (Compliance was high in both East Asia and Oceania.) It is interesting that East Asian countries differ from the Oceanic countries in respect of lacking island status (at least in the case of South Korea) and having high population density. Whatever differences explain the better outcomes in East Asia, it seems likely that they are features of the population rather than of any response strategy. This would explain why, for them, containment was much easier to achieve than for the Oceanic countries. And it would also explain why losing a grip on containment did not result in a rapid growth of cases, as it did in Melbourne.

The paradoxical upshot is that containment is a worthwhile strategy *only where* it is difficult and unstable, which is presumably only where success depends on features of the context (island status, low population density, and compliance) rather than features of the population (low obesity or other factors resulting in low transmission rates or mortality). If it is easy to achieve containment because of underlying population features that make the disease less likely to spread fast, then there is also less reason to seek to contain, since low levels can be maintained without it getting "out of control." Conversely, in a country where the population is highly susceptible, then strenuous efforts to achieve containment are bound to be required; and, if there is a realistic prospect of success, then strenuous efforts are also justified by the enormous benefits – which, however, are highly unstable, since if the disease does emerge in the population, it will quickly get "out of control."

Containment is thus a more complicated choice than it might at first appear. It is useful only for countries where it would make a big difference, but for the vast majority of such countries, it is unachievable. There remain only a few countries for which it is both achievable and worthwhile, and for these, it is difficult to achieve, unstable if achieved, and therefore, aiming for containment is a risky strategy. Very few states can claim to have successfully followed containment strategies during Covid-19, but many citizens of those states will be glad they did so. Containment becomes a more important strategy for less contagious diseases, however, and may prove more relevant in future pandemics where it is an option open to more countries. However, a less contagious disease is less likely to develop into a pandemic and more likely to be contained in the places where it first appears.

4.13 Conclusion

It might sound dramatic to claim that science caused the lockdown, but that is not far from being an imprecise statement of the truth. An intervention into public life by a scientific authority uniquely placed to influence global response to infectious disease outbreaks introduced an entirely novel distinction, which collapsed a universe of policy options into a binary choice. Prior thinking about layering different measures and weighing up risks and benefits went out of the window. The divergence between achieving suppression and failing to do so was so stark, it seemed, that doing the utmost to achieve suppression seemed, not just worth incurring high costs, but the only acceptable policy. That is a remarkable sequence of events, of which we are unaware of a precedent.

The next chapter contends that lockdown was the subject of a scientific dogma. Before turning to it, notice that methodological rigidity plays a connecting role here. Descriptive models would never have yielded and did not yield such a dramatic divergence between strategies. One reason for this is that the *R*-number is less central in descriptive models. The *R*-number is of central importance in mechanistic models because it quantifies how many other infected individuals will, on average, go on to infect, and this is a central determinant of the outcome of the model. But in descriptive models, it is not so crucial. It is something that is measured, rather than something that plays a causal role. Another reason that descriptive models would not have yielded a misguided policy binary is that actual outcomes did not diverge sharply and were not, even when ICL9 was produced. We saw this when discussing the IHME model, and we will see it again in the next chapter when discussing Sweden's approach. So methodological rigidity played a central role, not only in generating predictions, but also in shaping response strategies. And when these were not borne out by the evidence, they became the subject of a scientific dogma, to which we now turn.

Note

1 This chapter draws upon material in (Streicher and Broadbent 2023).

5

SCIENTIFIC DOGMA AND LOCKDOWN

5.1 Introduction

In this chapter, we introduce the idea of scientific dogma, and argue that the necessity and effectiveness of lockdown became the object of one. We demonstrate that these assertions were made or implied, and that there was resistance to acknowledging evidence to the contrary.

We begin by clarifying what we mean by "lockdown" and highlighting why it was so distinctive and why it was unprecedented as a public health intervention. We then introduce the notion of scientific dogma and propose that the claims of lockdown's necessity and effectiveness took on that status, being inappropriately maintained by scientific authorities in the face of recalcitrant evidence. To support this argument, we examine case studies of influential evaluations of lockdown that proved resistant to evidence challenging their conclusions. In particular, we show that early claims attributing large numbers of lives saved to lockdowns were based on initially undetected and later uncorrected data errors, and later preserved with modified methods (which nonetheless led to the estimates being modified downwards significantly, by approximately an order of magnitude). The necessity and effectiveness of lockdown resisted recalcitrant evidence. Their status as scientific dogma explains how a previously dismissed intervention came to be widely seen as a matter of scientific consensus, and widely implemented as such.

DOI: 10.4324/9781003732495-6

5.2 Lockdown

The most saliently unique feature of the Pandemic was the lockdown. Pandemics are not unprecedented, but lockdown was, on the scale seen during the Covid-19 Pandemic.

We define a lockdown as follows.

> A lockdown is a legal requirement for all persons to stay at home unless their reason for being outside the home falls within a closed list of exceptions.

A mere recommendation for the whole population to stay at home where possible, but without legal obligation, is not a lockdown. As the authors of one paper put it:

> In China, the United States, and European countries, governments have declared lockdowns when infections rose, issuing "orders" for people to stay at home. In contrast, unlike the lockdowns in these countries, restrictions during Japan's state of emergency had no legal binding force, and there were no fines, arrests, or other punishments imposed for leaving home during the state of emergency. The Japanese government only verbally "requested" people to refrain from going out, and many people heeded this "request" by staying at home.
>
> *(Watanabe and Yabu 2021, 334).*

Thus, Sweden, South Korea, and Japan had no lockdown, while European countries aside from Sweden did. A lockdown is thus a regulatory entity, whose effectiveness as a public health measure will vary depending on compliance, the number of people legitimately leaving the home under one of the exceptions, the nature of restrictions on activities outside the home (and compliance with those restrictions), and many other factors.

Lockdowns varied considerably in both their regulatory content and the changes in social contact that they achieved. Variation in both respects is possible. Two different regulatory packages might meet our definition of lockdown but differ markedly in what activities they permit (e.g. exercise outside the home; which occupations count as essential; etc.). And the same regulatory package implemented in two countries might result in different reductions in social contact due to differences in living conditions, population density, common occupations, compliance, enforcement, and so forth.

These variations are sometimes pointed to as evidence that there is no single such thing as lockdown. But that is unhelpful, since there is a common element to lockdown, which is the stay-at-home order, reversing the onus typical even in totalitarian states by requiring one to prove that one is justified in leaving the home.

The notion of a "de facto lockdown" was invoked at times to explain the absence of disaster in countries that did not lock down when their neighbours did (notably, Sweden). However, it is an incoherent notion. A lockdown requires a "lock." The key question around lockdown is not whether people should be *advised* to stay home, but whether they should be *required* to do so.

Many other public health measures exist short of lockdowns. These include: social distancing regulations, including limiting the capacity of restaurants, classrooms, buses, etc.; bans on large gatherings; school and university closures; case isolation; quarantining; contact tracing; applications of some of these measures to specific sections of the population (older people, etc.); and various possible uses of personal protective equipment (especially masks). There are, however, several important differences between lockdowns and other public health measures, including others that were controversial (such as school closures and mask mandates). Here are four.

First, other measures can be *layered*, while lockdown is hard to layer with other measures. With a lockdown in place, most of society is affected: buses, if they run, will be empty; schools will be closed; and so forth. It is true that lockdowns can vary in their strictness because the list of exceptions can vary. And a sufficiently huge list of exceptions could, in principle, result in something little different from a stringent package of non-lockdown restrictions. But would the result really be a lockdown, in anyone's eyes? The concept of a lockdown, and the spirit of it, is that you do not go out unless you absolutely have to. This reversal of the usual burden of reasons for leaving the home is fundamental, and it is a categorical matter. Once that is in place, and assuming it is not eroded by a panoply of exceptions, there is little scope for layering other public health measures, since most aspects of society will already be massively affected by the lockdown.

Second, the social, economic, political, and personal *consequences* of a lockdown are far greater in degree and extent than those of any other public health intervention. During a lockdown, fundamental social and personal activities become impossible for practically everybody, outside of limited exceptions. States differed in the range and rationality of their exceptions, but in all cases, this situation placed a great burden, not only on individuals (as is commonly pointed out), but also on many societal structures, entities, and activities. Businesses were the most obvious, but business is just a formalized aspect of the complexity and richness of any society of humans. The effects of lockdown include those that are obviously negative: loss of ability to work; loss of livelihood (e.g. business, livestock); inability to access healthcare; loss of education; exposure to domestic abuse; damage to mental health; for many, intolerable living conditions due to crowding and heat; and others. One could restrict access to healthcare, close schools, prevent certain sorts of economic activity, and so forth, without locking down, but a lockdown implies that all these things happen at once. Therefore, the negative effects of lockdown are necessarily much larger and broader than those of other interventions, providing an excellent reason to single lockdown out for special consideration.

Third, a lockdown requires a *violation of rights* of a kind that is at least quantitatively and perhaps qualitatively different from that required by the other measures. The right to liberty is the obvious one. Lockdowns reverse the burden of justification for carrying out any activity in a very large domain, that is, outside the home. Usually, individuals may do what they wish with certain exceptions that are specified by law. Citizens of totalitarian states and persons persecuted in certain jurisdictions do not enjoy this benefit. However, most of the world's population is generally able to carry on as it likes, provided that it does not transgress certain rules, whether or not these are fair, clear, applied consistently, etc. However, lockdown reverses this situation. Doing anything outside the home requires proof that the activity falls into one of a number of exceptions to the stay-at-home order. Even in countries where liberty rights are not well-protected (e.g. where police commonly harass people), a lockdown is a very significant reduction of liberty. Note, however, that liberty is not the only right affected. Others include the right to access healthcare, the right to work, and freedom of association.

Fourth and finally, the use made of lockdowns during the Pandemic is *unprecedented*. Lockdowns happened in China, but were not considered feasible in Europe prior to Italy implementing one. In an interview with British newspaper *The Times*, the leader of the ICL Covid-19 Response Team said:

> [China is] a communist one-party state, we said. We couldn't get away with it in Europe, we thought . . . and then Italy did it. And we realised we could.
> *(Sayers 2020)*

Stay-at-home orders had been implemented in various circumstances previously, but never on a scale affecting millions of people, by nation-states, let alone nations in the historical heart of the nationalist project – Europe. This is a generalization that applies to *any* kind of lockdown, not just a health-related one, making the uniqueness two dimensional.

Not only were lockdowns previously rare, but they were also contrary to planned responses to a viral pandemic, which emphasized "minimising potential impact of a pandemic on society and the economy," "supporting the continuation of everyday activities as far as possible," and "promoting a return to normality and the restoration of disrupted services at the earliest opportunity" (DH Pandemic Influenza Preparedness Team 2011). The UK's plan was widely regarded as exemplary, and the content of the WHO's 2019 plan was essentially the same (WHO 2019). The words "lockdown" and "suppression" were not used in pre-pandemic plans. Aside from minimizing social and economic impacts, which was obviously not the theme of lockdown rhetoric in the pandemic, regulatory approaches were thought to be unnecessary, unfeasible, or both (Inglesby et al. 2006).

5.3 Scientific Dogma

In one dictionary definition, a dogma is a principle or set of principles laid down by an authority as incontrovertibly true.[1] Others are similar. This definition indicates that a dogma is a proposition with a certain social significance, rather like an insult. The propositional properties of a dogma include its assertoric nature and its content. The non-propositional properties concern its source, "authority," and the power of that source to "lay it down."

For our purposes, a more focused definition is necessary, one that preserves the usual sense but clarifies the source of authority and identifies the characteristic form that dogma takes in science. Unlike the medieval church, science does not usually state "incontrovertible truths." Indeed, "truth" is not commonly used in scientific discourse at all. "Fact" is popular in rhetoric and discourse *around* science, but less common within it. A more natural expression of scientific dogma might involve the term "evidence:" science is centrally concerned with evidence, and often distinguishes itself from religion and other epistemic endeavours by this concept. So it would be natural for scientific *dogma* to include an unusual attitude towards evidence.

We, therefore, propose the following.

A scientific dogma is

(a) A proposition
(b) Asserted by a scientific authority
(c) And inappropriately maintained in the face of recalcitrant evidence

To show that (a) is satisfied requires identifying the content of a dogma. This is not always easy since a scientific dogma is obviously not presented as such, and thus does not receive a clear formulation, as a religious dogma usually would. We must usually infer the content of a scientific dogma from its defence. It cannot be expected that this content will be very precise, and it cannot be required of a scientific dogma that it is never contradicted by explicit statements of those whose other actions and statements show that they endorse it.

To show that (b) is satisfied requires identifying an assertion of the dogma, and thus also identifying the assertor as a scientific authority. Since scientific dogma is not explicit, as just discussed, the notion of assertion in (b) cannot be confined to uttering the proposition. We must also allow a proposition to be asserted in other ways, for instance, by defending it, or insisting that those who deny it or parts of it are wrong. Although philosophers are accustomed to using "assertion" as indicating the introduction of some symbol into discourse (uttering some words, writing something, etc.), the usual sense of the word is much broader, and includes phrases like "asserting one's authority" and "asserting oneself." It is this broader sense of "assertion" that (b) invokes, which does not commit the assertor to uttering the content of the dogma.

As to the assertor's identity as a scientific authority, it is usually easy enough to decide whether a person or institution is a scientific authority of some kind. One possible objection here concerns showing that the dogma originates with that authority, rather than simply being parroted. This is especially so when there is no clear origin, or when the origin story of a scientific dogma involves multiple actors – as is the case with lockdown, as we shall see.

However, our definition is satisfied so long as a scientific authority does as we describe. It makes no mention of whether other scientific authorities do so too, nor of who asserted it first. All it takes for a scientific dogma to exist is for a scientific authority to assert it. This implies, of course, that scientific dogma need not represent a consensus of scientific authorities, and indeed might be highly localized – even so far as being asserted by just one authority. This is all consistent with our aims, since we are clear that the lockdown dogma (as well as the others discussed in the following chapters) did not represent anything like a scientific consensus, and probably represented the view of a scientific minority.

The satisfaction of (c) requires showing that the content of the dogma is not merely asserted as fact, but inappropriately maintained in the face of recalcitrant evidence. This is a matter of degree, making this a natural point for objection. Surely many scientific assertions are at least *slightly* resistant to recalcitrant evidence? If children boil a flask of water in class, and observe that the boiling point is a little under 100 degrees Celsius, the teacher will explain to them that this is because there are substances dissolved in the water that reduce its boiling point. The teacher is asserting a proposition – that water boils at 100 degrees Celsius – in the face of recalcitrant evidence – a classroom full of beakers bubbling at a slightly lower temperature. Is the teacher therefore being dogmatic, on our definition?

Our response is that the semantics of "dogma," as of most natural language terms, are contextual. In context, the teacher is not being dogmatic because it is perfectly appropriate to maintain the proposition in this context. The teacher is conveying some building blocks of scientific thinking to young minds. These include: that liquids have characteristic boiling points across a broad range of conditions; that water's boiling point is around 100 degrees Celsius; that substances dissolved in water can lower this; and that the degree – the unit of measurement of temperature – was historically calibrated using the freezing and boiling points of water. In that context, the teacher's assertions are not dogmatic because the purpose of the assertion is served and not vitiated by persisting with the claim in the face of recalcitrant evidence.

On the other hand, in a stricter sense, the teacher is not being accurate. Hasok Chang has famously argued against the idea of a fixed definitional boiling point of water at 100 degrees Celsius. The reality is more complicated. Boiling is not a straightforward phenomenon, and "Water boils at 100 degrees Celsius" is, from a certain perspective, little more than a local approximation. So, there are contexts in which it would be dogmatic to insist that water boils at 100 degrees Celsius. At a physics or philosophy conference, it could well be dogmatic to assert that water boils at this point. It could be dogmatic in a classroom, too, if an intelligent and well-informed student asks good questions and is shut down. Whether or not the resistance to recalcitrant evidence is inappropriate thus depends on context.

We see this as broadly in tune with general use of the term "dogmatic," which permits degree and contextual variation. We do not believe it would be helpful, either conceptually or practically, to identify a sharp line between dogma and non-dogma. In practice, certain things are a matter of judgment, and overtheorizing about contexts one is not actually in can amount to a kind of hubris.

In the context of the Pandemic, we shall argue that propositions were asserted by scientific authorities and inappropriately maintained in the face of recalcitrant evidence – inappropriate, because doing so vitiated the purpose for which the assertions were made. This is the right level of theoretical detail to allow us to proceed.

5.4 The Lockdown Dogma

We say that there was a *Lockdown Dogma,* as follows.

> **Lockdown Dogma**
>
> Lockdowns, properly executed, were widely necessary and effective for responding to epidemic waves in the Pandemic.

The Lockdown Dogma makes space for poorly timed lockdowns, badly implemented lockdowns, and even overzealous lockdowns. The definition above is not intended as – indeed, cannot be – strict, for reasons discussed already: scientific dogmas, in contrast to religious ones, are not explicitly stated as such. However, this definition, loose as it is, captures an important and recognizable sentiment that was widespread and powerful during the Pandemic.

An important point about the content of this dogma is that necessity and effectiveness are not the same thing. Lockdowns must have some effect to be necessary, of course. But they can have an effect even if they are not necessary, since necessity is relative to some objective. During the Pandemic, the standard objective had something to do with the healthcare system, represented in ICL9 as a straight line, depressingly low on a graph of mountainous curves. However, as we shall see in the next chapter, subsequent evaluations of lockdown did not make any reference to healthcare capacity at all. They simply sought to show that lockdowns had had an effect, usually a large one, and that they had brought about a decline in infections.

We now conduct two case studies in which *the necessity and effectiveness of lockdowns were inappropriately maintained in the face of recalcitrant evidence*. Both evaluations were conducted by the ICL team in publications of June 2020 and August 2021 respectively. Both maintained that lockdown was necessary and effective, as we shall explain, in the face of evidence to the contrary. This "maintaining" was more sophisticated than simply digging in heels and insisting. We will identify a variety of features of these studies which, we will argue, amount to maintaining in the face of recalcitrant evidence. We will not spend time arguing that this was inappropriate because it is obvious that the context was one in which the necessity and effectiveness of lockdown were precisely at issue. The assertions about the effects of lockdown were being made to inform the public as to whether those lockdowns had been necessary and effective, and possibly also to inform future policy. Unlike the teacher glossing over the details of the boiling point of water, scientists erroneously (as we shall argue) saying that lockdowns saved millions of lives cannot reasonably be excused from their error by appeal to context.

In arguing for the inappropriate treatment of recalcitrant evidence, we do not, strictly speaking, need to establish that the evidence is incompatible with the treatment we criticize. More concretely, in order to criticize a specific, positive evaluation of a particular lockdown (or group of lockdowns), we do not need to show that these lockdowns were *unnecessary* and *ineffective*. We need merely to show that *the evidence was not conclusive in the way it was supposed to be*. However, the more strongly the evidence points away from the conclusion that was drawn from it, the stronger our case will be, and this naturally results from showing how the evidence favours an alternative hypothesis. In particular, identifying evidence that *differentiates* between two hypotheses, being explained by one but not the other, is compelling (Lipton 2004, Chapter 5). Thus, we will argue that lockdowns were ineffective or unnecessary in the two instances we discuss. However, this is not strictly necessary for our conclusion; it is enough to accept a weaker stance, that the evidence was not handled impartially in the studies we criticize.

5.5 Did the UK Lockdown Save 470,000 Lives?

On 8 June 2020, the BBC reported that lockdowns saved 3.1 million lives in Europe and 470,000 lives in the UK (Gallagher 2020). This was based on a press release from ICL on the same day with the title: "Lockdown and school closures in Europe may have prevented 3.1m deaths," including this sentence: "it is clear to us that non-pharmaceutical interventions – such as lockdown and school closures, have saved about 3.1 million lives in these countries" (Wighton 2020).

The ICL press release promoted an ICL paper published in *Nature*, also dated 8 June 2020, with the title: "Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe" (Flaxman et al. 2020a), from here on referred to as Flaxman et al.

For many, this BBC report had a lasting effect on what people believed about the effect of lockdowns. The claim of 470,000 lives saved was repeated in the UK Parliament as late as 18 January 2022 by the MP, Brendan O'Hara: "A report by researchers for the journal Nature found that the first lockdown saved up to 3 million lives in Europe, including 470,000 in the UK" (Hansard 2022).

Yet, by 23 December 2020, the authors of the paper in question had already significantly toned down their claims, stating that "Although our work shows that lockdowns had the largest effect size, we did not and do not claim that they were the only path to controlling the virus; merely that among the NPIs we considered, lockdown was the most effective single measure" (Flaxman et al. 2020b), and moreover that it was the only one with an identifiable effect (the remaining effects being small and cumulative). However, graphically, the response indicated that in Sweden (which did not lock down), the banning of public events had as large an effect as that of lockdowns in most other European countries. The same representation showed that the banning of public events had a small effect in most other European countries. No comment accompanied these graphs. It is obvious that something has gone wrong – but what?

5.5.1 Lucky Sweden

The paper prompting the aforementioned reply came from a Swedish team, which pointed out that lockdown was not the most effective measure in one of the 11 countries considered: theirs (Soltesz et al. 2020). Sweden did not lock down, but R_t fell below 1 in roughly the same way as in other European countries. The Flaxman et al. model accommodated this by including random variation, which, in the case of Sweden, was sufficient to account for R falling below 1. Effectively, Flaxman et al. said that in ten countries, R_t fell below 1 because of lockdown in ten cases and by luck in the eleventh. It must have done, since there was no lockdown. If this was the reasoning, it is obviously circular; if this is not the reasoning, it is unclear what was.

ICL replied that the point was "interesting" but that their goal had never been to explain what happened in any particular country, only to identify what worked in general:

> The main goal of our paper was to examine multiple countries to see what worked in most places, not to explain the trajectory of the epidemic in each individual country.
>
> *(Flaxman et al. 2020b)*

And:

> our goal was to estimate which NPIs worked consistently in most countries, we argue that an analysis of the effectiveness of NPIs should be robust to leaving any one country out.
>
> *(Flaxman et al. 2020b)*

Thus:

> within our model, we can conclude that all NPIs together brought the epidemic under control; lockdown had the strongest effect and was identifiable; and the precise effect of the other NPIs was not identifiable.
>
> *(Flaxman et al. 2020b)*

This, however, is not the same thing as saying that lockdowns had a strong, identifiable effect in all countries, which was what the original paper claimed. They may have had a major effect in all countries *where they were implemented*, and, if that is a large majority of countries, it supports one interpretation of "lockdown had the largest effect." But that ambiguous phrase could also be interpreted to say that lockdown occurred everywhere, which it did not, and so that interpretation must be discarded. A third interpretation is that lockdown had a larger effect than any other approach, *including* the approach of the one country that did not lock down. This is impossible to justify in this paper, but it was the interpretation that the rest of the world adopted.

In their reply, the authors emphasize that there are other paths besides lockdown to bringing R_t below 1 (Flaxman et al. 2020b). However, this is definitely not stated in the original paper. The British media interpreted that paper as showing that lockdown had saved 470,000 lives – something that stretches well beyond anything that the paper could possibly have shown, even on its own terms, since the identifiable effect of lockdown was relative to previously applied interventions. At best, the paper was irresponsible in presenting its findings in a way that was open to quite a different interpretation than could be justified.

5.6 Corrected Data Imply Peak Occurred Before Lockdown

The time between infection and death was wrongly calculated, and when that error is corrected, the paper's own model places the peak of infections in England earlier than the lockdowns that were supposed to have caused it. If so, it is impossible that lockdowns caused the peak in England.

The error arose because the time of *report* of death was used as time of death itself for the UK. Figure 5.1 compares official daily deaths data with that used in the paper under discussion. There is a time lag between the two, with deaths occurring earlier in the official data than in the paper's data. This is an error in favour of the paper's conclusion. An additional piece of evidence that the data relied on by the paper are erroneous is a "weekend effect." Reports of death fall off at weekends. Covid-19 does not observe weekends, however.

The consequences of this error are severe. Infections are inferred from the date of death, which is inevitably somewhat speculative, since this interval must be estimated. The paper uses an interval of 22.9 days on average between infection and death, which is within the range of reasonable estimates, albeit on the long side. Even on this basis, lockdown occurs on the day of peak infections, which would be a little surprising – one would expect infections to peak at least a day or two later because people would still be infecting others in their homes. However,

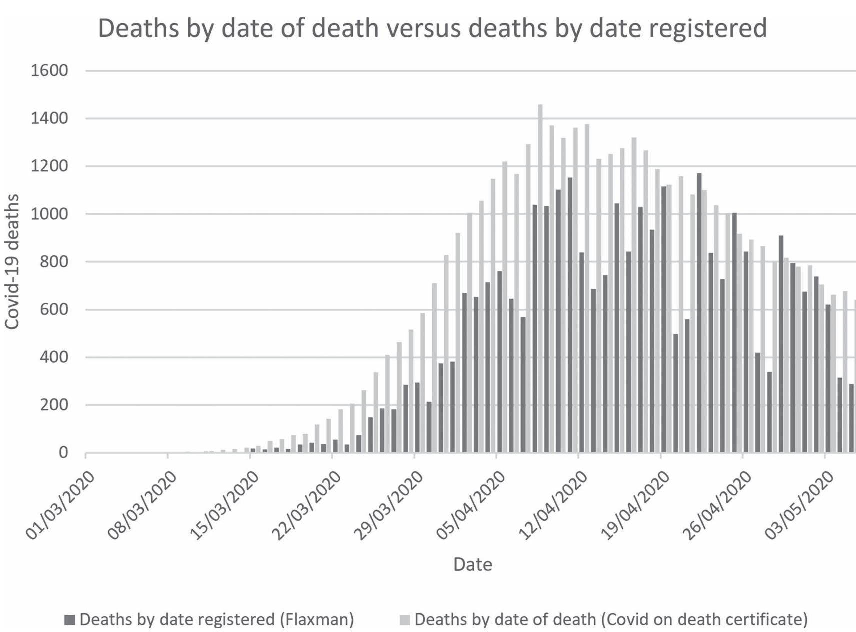


FIGURE 5.1 UK deaths per official data versus Flaxman et al. 2020a

when we add 3–5 days to the interval, peak infections occur *before* lockdown. Thus lockdown could not have possibly caused the peak. It was, therefore, neither necessary nor effective at bringing *R* below 1, according to this paper. The error, however, supported the paper's conclusion, which was in this way resistant to recalcitrant evidence.

Such an error ought to have been picked up by a sensitivity analysis. A peer reviewer requested this. The analysis was duly done, but only in one direction: that is, for a shorter latency period. The effect of a longer latency period was not considered. Obviously, a longer latency period would undermine and, at a certain point, entirely refute the conclusion. Again, the paper evinces insensitivity to recalcitrant evidence.

Here is the reviewer's request, and the response.

> [Reviewer:] Page 3, last paragraph – a citation justifying this 2–3 infection acquisition to death time lag is needed, presumably based off patient line-list data.
>
> [Authors' response:] We have added a citation – we use the onset to death distribution specified by Verity et al. 2020, with a mean of 17.8 days. However, we do acknowledge that other values exist and have done a sensitivity analysis running our model using an onset to death mean of 15 days (Khalil et al 2020) and for 13 days (Linton 2020 and Deng et al 2020). Changing this does increase the number of deaths and hence deaths averted, but not outside the credible intervals of our chosen value of 17.8.
>
> *(Anon., n.d.)*

Thus, the possibility was considered that a shorter latency period would result in effectiveness too great to be credible, but not whether a longer latency period would result in effectiveness too small to be worthwhile, or present only after the peak.

It could be that voluntary behaviour change explains why the peak occurred before lockdowns were implemented. In that case, Flaxman et al show that voluntary behaviour change can bring about a peak, and that lockdowns are not necessary. The paper also finds no effect of the actual lockdown, over and above what was already being achieved through voluntary behaviour change. This is ridiculous. Life under lockdown was radically different from life before, even if there were some modifications of behaviour. Nobody living through that period can accept this response.

5.7 Ecological Fallacy

Flaxman et al. commit a causal inference error by assuming that a change in R_t can only be attributed to the previous intervention. This is the simplest causal fallacy, variously known as the ecological fallacy, the phrase "post hoc ergo propter hoc," and the slogan "correlation is not causation."

One might defend it as a simplifying assumption. However, it is not an appropriate assumption because of the fact that R_t falls from early on in an epidemic wave. It was obviously falling at the time of the final intervention (lockdown) since the paper considers previous interventions and finds each reduces R_t. Therefore, one cannot infer that the intervention caused R_t to fall below 1, without as a minimum indicating what R_t would have been without the intervention, for example, through a curve-fitting exercise. In effect, Flaxman et al. perform such an exercise, but extremely poorly, projecting a straight line, which is so obviously a poor fit for the previous data that it escapes notice as a curve-fitting exercise at all.

Aside from R_t falling between interventions, the causal assumption fails to take account of the numerous potential confounders, such as seasonality and immunity. One can assume that seasonality has a negligible effect within this short timeframe, but one should state and justify the assumption since a seasonality effect is commonly observable with respiratory viruses, such as influenza (it was observed with Covid-19 in some places, but not in others). Immunity must also be considered. This would also have forced the modellers to consider the falling value of *R*, since one reason that *R* falls is that the agent runs out of susceptible individuals, and immunity to re-infection is one reason for that.

5.8 Circularity

Combined with the causal assumption, the assumption that *R* does not change between interventions results in circularity: the method effectively guarantees its conclusions, under certain easily satisfied conditions. Any intervention occurring after *R* starts to fall will have its effect overestimated on the assumption (shared by Flaxman et al.) that there are no endogenous events, and that *R* will fall between any two times t and $t+1$ (within a reasonable interval of interest). When *R* falls below *1,* this means that the last intervention while *R* was above 1 is credited with causing *R* to fall below 1. On the paper's assumptions, the final intervention is guaranteed to be credited with causing *R* to fall below 1, provided it satisfies the following conditions.

The Flaxman et al. Test for Necessary and Effective "Suppression"

1 When intervention *I* occurs at t, $R_t > 1$.
2 No interventions occurred between t and t+1.
3 $R_{t+1} < 1$.

For convenience, we use the term “suppression” here in the sense we have rejected, that of causing R_t to fall below 1. Due to its implicit causal assumptions, the Flaxman et al. method will show that any I has the effect of reducing R_t to R_{t+1}, provided no other interventions occurred in the interim, as provided by (2) (it does not matter whether the estimate at $t+1$ is accompanied by a further intervention or not). If $R_{t+1} < 1$, as provided by (3), then I caused the fall.

Flaxman et al. believed that the last intervention before the peak was a lockdown, and therefore concluded that the lockdown caused the peak. However, the reasoning goes equally if I was the banning of the sale of garden hoses.

Where lockdown restrictions were lifted while R_t was falling but still greater than 1, the same method can be used to show that lifting restrictions caused a reduction in R and eventually caused R to fall below 1. In South Africa, R_t was above 1 but falling when Alert Level 5 – a very strict lockdown – was slightly reduced to Alert Level 4. This was still a very severe set of restrictions, more stringent than many lockdowns in other countries. Cases were still rising when Alert Level 4 gave way to 3, but they peaked during 3 (Broadbent, Combrink, and Smart 2020). Therefore, lifting restrictions from Level 5 to 4 improved matters somewhat, but lifting them from 4 to 3 caused R_t to fall below 1. In reality, there was little noticeable change in the rate of change of R_t on either occasion, but if there was a change, it was certainly not large enough to push R_t upwards again. Thus, lifting restrictions proved an effective suppression method in South Africa. This is a reductio ad absurdum.

5.9 Forgetting the Target

Aside from these points about the insensitivity of the paper to the evidence, another important feature that this paper shares with many others is that it completely ignores the question of whether lockdowns were *necessary*, and focuses solely on whether they were *effective*. This is the point Woolhouse makes in his testimony to the UK Covid-19 Inquiry (UK Covid-19 Inquiry 2023b). Almost every prospective argument for lockdown in the UK depended on “protecting the NHS” (i.e. the National Health Service), which was cleanly represented as bringing a curve below a flat horizontal line representing ICU bed capacity. The IHME also focused on healthcare capacity. So did Sweden and South Africa. ICL brought out a further report concerning the whole world, also relying heavily on the extent to which hospital capacities would be exceeded in calculating mortality (Walker et al. 2020). When motivating lockdowns, their necessity was relative not simply to causing R to fall below 1, but to protecting health services from the consequences of failing to do so.

Therefore, an evaluation of the success of lockdown in a country such as the UK, where this was a paramount motivation, needs to consider whether it was necessary for protecting the health service.

Yet, there is no mention in the paper under discussion of whether lockdowns were necessary for keeping severe infections below this threshold. The point is simply not discussed. This is consistent, of course, with the assumption that R_t does not change between interventions, since in that case without intervention infections will rise to infinity, which is beyond the capacity even of the NHS.

The point re-emerged after the Pandemic. Three years later, the leader of the team at ICL (and number 51 of the 53 authors listed on the paper in question) testified as follows to the UK Covid-19 Inquiry.

Lead Inquiry: In your view, if the goal was to prevent the collapse of the NHS, was that lockdown [in the UK, 23 March 2020] necessary?

Professor Neil Ferguson: This is not a question we can definitively answer. Without doubt, the measures announced on 16 March had some effect on transmission, and potentially accelerated by spontaneous behaviour change, but we didn't have the time to wait and collect the data which would allow us to say, "Yes, they're sufficient," or "No, they're not."

Lead Inquiry: My question to you was: if the goal was to prevent the collapse of the NHS, was that lockdown necessary? From everything you've said, it must surely follow that it was.

Professor Neil Ferguson: Okay. So I think – you know, well, I think both were warranted, but I cannot definitively say whether what was announced on the 16th, maybe in combination with what was defined – announced on the 18th, would have been sufficient in its own right, we just don't have the data to answer that question. (UK Covid-19 Inquiry 2023c)

If the data to answer this question were not available when Ferguson gave evidence on 17 October 2023, then it was not available when Flaxman et al. was published in June 2020.

5.10 Did the UK Lockdown Save 40,000 Lives?

In August 2021, the ICL team published a paper comparing the UK, Denmark, and Sweden, with a view to assessing how their respective policies contributed to their outcomes (Mishra et al. 2021). Despite being published in 2021, it only considered the period up until July 2020. Like the previous paper discussed, it was published in a major journal, although it attracted less press attention and fewer citations. This could have been due to timing. It could also have been due to the fact that the number of lives saved in the UK was now 40,000, down from 470,000 in Flaxman et al. (2020a). This was because the comparator was no longer a counterfactual do-nothing scenario, but a counterfactual scenario in which a different set of measures were implemented in the UK. This is much more reasonable as a way to express effectiveness, although it implies that 91%

of the benefit attributed to lockdown on the basis of Flaxman et al. could have been achieved without it. The paper also corrects the data error in Flaxman et al. Nevertheless, it remains highly resistant to recalcitrant evidence.

Before considering its merits, it must be noted that the paper starts with a claim that was already false at the time. The first sentence of the abstract states:

> The UK and Sweden have among the worst per-capita COVID-19 mortality in Europe. Sweden stands out for its greater reliance on voluntary, rather than mandatory, control measures.
>
> *(Mishra et al. 2021)*

This was true neither on 31 March 2021, when the paper was submitted, nor on 27 July 2021, when it was published. Covid-19 mortality per million for the three countries at the latter time was, respectively, Denmark: 422; Sweden: 1302; and the UK: 2268 (Our World in Data 2024). It is clear that mortality per million is much worse in Sweden than Denmark, but equally clear that it is much better in Sweden than the UK, which, by this time, had locked down in response to the second wave as well as the first. So, the entire rationale for the selection is undercut. Sweden was between the other two.

It was, moreover, no longer the case that Sweden was among the worst performers in Europe; it was, on the contrary, in the better half. This is not to vindicate Sweden's policy: it is merely to point out that this paper is premised on a falsehood, which it nowhere acknowledges. And subsequent events made matters worse: by the end of the Pandemic, Sweden would be in the top third (again, a statement of fact; no evaluation of Swedish strategy implied).

The issue arises because the paper considers only the first wave. The statement is more defensible at that point in time (although "stands out" remains objectionably subjective). But time-indexing provides no defence, because the paper sought to establish conclusions about how different measures and their timings impacted the "final death toll," not about the death toll at the end of the first wave. To say that a country *has* a poor per-capita mortality and that it *had* one a year ago are different; the former statement is the one that was made, while the latter is the one that was evaluated. This is all the more egregious given that Sweden's strategy was in part motivated by an eye on the long term.

The choice of countries can only be understood as politically motivated, at least in part, in a way that damages the epistemic properties of the study. In the context of the time, when Sweden was heavily criticized by a substantial portion of society, there is no way not to see this as an effort to assess Sweden's policy. As for the UK, the team is obviously based in the UK, advising UK policy, and the British government had likewise faced criticism for acting too slowly. Combined with the previous point – that, at the time of publication, the proffered rationale for picking these three countries no longer held true – it is absolutely reasonable to infer that the motivation for the selection was, at least in part, political and was not purely epistemic.

5.11 Critiquing the Revised Estimate

We will be briefer with Mishra et al. than with Flaxman et al. because the pattern of issues is similar. The paper fails to answer an obvious question about Sweden, just as the previous did (albeit a different question); it corrects the data error from the previous paper but uses different methods to arrive at a similar – albeit smaller – positive result; it does not evaluate lockdown against the stated policy goal of staying within hospital bed capacity; and it carries forward the methodological errors of its predecessor. The reader who does not wish to rehearse these points can skip the remainder of this section.

5.11.1 Unanswered Question

The paper does not fully answer the tasks it sets itself. One is to assess the relative effectiveness of the *nature* of the NPIs in the three countries: that is, the relative effect of lockdown and the Swedish approach. The other is to assess the relative impact of the *timing* of the introduction of the measures. Denmark locked down early. Sweden implemented non-lockdown NPIs early. The UK locked down late. So, the paper asked what would have happened in Sweden if it had locked down at the same early time that it implemented its non-lockdown NPIs, and what would have happened in the UK if it had locked down at the time Denmark did. It also considers what would have happened in the UK if it had implemented the Swedish approach at the late time that it in fact did implement a lockdown. But it fails to consider what would have happened in the UK if it had done what Sweden did, that is, implement non-lockdown NPIs early. So, it fails to fully answer the question it sets itself.

This is despite the interim conclusion that the major difference between Swedish and Danish approaches was not the eventual extent of behavioural change and social contact reduction (eventually similar between the two countries), but the time it took for behaviour to change (being slower in Sweden and abrupt in Denmark) – implying that an earlier introduction of suppression measures short of lockdown might have achieved adequate contact reduction in good time, rendering lockdown superfluous. Or perhaps not; the study does not answer this crucial question.

This is important because lockdown is a policy that one does not follow unnecessarily.

During the UK Covid-19 Inquiry, several scientists testified that an earlier introduction of NPIs short of lockdown might well have rendered lockdown unnecessary. At the time, uncertainty might have meant they could be justified as a precaution. But here we are talking about retrospective evaluation of effectiveness and not about justification. Evidently, this evaluation will inform subsequent policy decisions. Given the undesirability of lockdown, it is extremely important, perhaps of paramount importance, to ask whether an earlier introduction of lesser methods could have done the job. Instead, the paper asks all other possible questions in this logical space and omits that one: the one whose answer might contradict the Lockdown Dogma.

5.11.2 Correcting the Data but Adjusting the Method

Mishra et al. corrects the data error in Flaxman et al. That is, it uses the date of death rather than the date of report. However, it reaches the conclusion that infections peaked after lockdown. This is because it uses a different method. Mishra was the second author on Flaxman et al. Flaxman was the fourth author on Mishra et al. There is significant overlap between the author teams. The earlier paper, Flaxman et al., is cited in support of the claim that "Denmark introduced major social distancing policies a few days earlier than both of the other two countries," but otherwise not discussed. It seems that the team was aware of the previous data error, and presumably of the fact that correcting it undermined the conclusion of that paper. The failure to say so cannot be condoned, since published work may be relied upon, and authors have a duty to correct errors they become aware of.

The use of a new methodology on the corrected data is more important from the point of view of establishing resistance to recalcitrant evidence. Obviously, if one adjusts one's method in light of new data, one is open to criticism. Indeed, when assessing pharmaceutical interventions, pre-registration of trial protocols is required to prevent this. No such restrictions exist for non-pharmaceutical interventions because these are not subject to commercial profit incentives. However, incentives operate in science nonetheless, including the incentive of preserving one's previous conclusions from contradiction and, in some cases, defending a dogma, as well as protecting one's indirect political authority and preserving dogma – further elements of scientific orthodoxy. This adjustment of methods is another fragment in the body of evidence that precisely this was occurring.

5.11.3 Hospital Bed Capacity Not Reconsidered

There is no reconsideration of hospital bed capacity, only eventual mortality. All three countries took their respective decisions with reference to whether hospital beds would be overrun, as discussed already. The Swedish approach succeeded *against this measure*, as discussed in the previous chapter (which, to repeat, is not to say that the Swedish strategy was an overall success, that it was optimal, etc.). That approach was to model likely hospital bed demand based on Wuhan's experience, and these projections proved to be good guides. The UK's experience was far more chaotic, and the fact that the NHS coped is less easily credited to good planning. The exertions of health workers played a major role. Nonetheless, in the end, the NHS did cope; it was not overwhelmed. If that was the objective, both countries' policies succeeded, or at least did not fail. It may still be worthwhile to consider eventual mortality, but to omit the question of hospital bed capacity is to sell all three countries short. It is also to ignore a potential comparator of relative success, and a relevant rationale for policy choice.

5.11.4 The Methodological Errors Are Carried Forward

Like Flaxman et al., Mishra et al. once again made unrealistic assumptions about R_t. In this model, R_t continues unchanged between interventions until two-week elapse, at which point, the new actual value of R_t is inserted. This again means that excessive effects are credited to each intervention. This does not create quite the same circularity as previously, since the goal is to compare countries rather than assess lockdown effectiveness directly. It does, however, constitute insensitivity to the underlying reality of the respective epidemic waves in a way that overestimates the effectiveness of interventions.

Also, the methodology is once again too insensitive to confounding differences between countries to support causal inferences about the relative effects of their policies. It may be an interesting exercise in counterfactual modelling to seek to counterfactually apply the strategy of one country to another. It may be valuable as a mathematical exploration of a possible technique for application in the future. But it *obviously* ignores differences between countries, and is therefore *obviously* unsuitable for drawing causal inferences. The paper is not an empirical study: it is a thought experiment.

5.12 Conclusion

Lockdown was the most striking feature of the Pandemic, and the one most deserving of terms like "unprecedented" and "unique." An implication of our treatment is that it was a product of scientific intervention into policy. Scientific opinion shaped the conceptual framework in which lockdown made sense. As we discussed in Chapter 4, although the word "lockdown" itself was not introduced by Pandemic-era science, the term "suppression" was. Because of their central role in shaping the intervention, the group that did so was motivated to assess its success rather than assess the success of the predictive models. The evidence was then interpreted as supportive when in fact it was not, first due to a data error, which was pointed out at the time but not corrected publicly, and then due to an adjusted method. The necessity and effectiveness of lockdown thus took on the status of dogma, asserted by a scientific authority and inappropriately maintained in the face of recalcitrant evidence.

Lockdown was not the only object of dogma, but it set a tone that made other dogmas more likely. We turn now to these.

Note

1 This is the definition Google offers at time of writing, based on Oxford Languages.

6

SCIENTIFIC DOGMA AND MASKS

6.1 Introduction

Mask use among medical professionals is commonplace, and in Asia, widespread mask-wearing has been popular for a long time. However, public health strategy has not previously considered general population masking to be an effective or necessary part of pandemic response. Yet during the Pandemic, masks became symbolic in the West, and particularly in America, almost like a flag worn on the face. This was out of all proportion to both their public health significance and their significance for liberty or any other rights, since masks are neither massively effective nor a massive imposition on the rights of the individual. A scientific dogma emerged around masks, and, while that dogma surely cannot be blamed for the entirety of the polarization, it certainly occupied one pole. In this chapter, however, we do not comment on the politics of masks, but on the scientific dynamics around their necessity and effectiveness. One's position on masks depends largely on one's position in relation to different types of evidence: that is, how sympathetic one is to observational evidence at critical risk of bias, as against how insistent one is on evidence from randomized trials before drawing a causal inference. This is an old debate and one that cannot be easily resolved. Evidence for the existence of a Mask Dogma comes from the failure to acknowledge the diversity of evidence, as we shall see. Masks represent an excellent, contained case study of the potential for conflict between observational and randomized studies, as well as some examples of how not to resolve it.

DOI: 10.4324/9781003732495-7

6.2 Masks: Pros and Cons

Masks have many desirable features as a public health measure. They are cheap to make and easy to distribute. They can be implemented flexibly at the population level or to sub-populations. Mode of implementation is likewise flexible, ranging from recommendations to full-fledged mandates. They rely on individual action, which is a merit in a context that prioritizes consent, and they offer benefits to the individual, which not every public health measure does. Masks might be thought to permit something closer to normalcy than some other measures, since they permit people to interact physically to a greater extent. (Conversely, if people are not interacting physically, masks are pointless.) The human cost of masks (harm in all dimensions) is small. Finally, their mechanism of operation is – on the face of it – easy to understand, which greatly supports compliance.

On the other hand, masks suffer some drawbacks. Individual control is one of them in a context that emphasizes compliance: a population-level effect will be smaller or absent if significant numbers of people do not wear masks. Masks need to be worn and disposed of correctly, changed often, and so forth, all important matters that masks place in the hands of the individual. Non-obvious harms might exist: for instance, it was hypothesized that children of some ages might suffer developmental issues around language or emotional recognition. Some people might have psychological or physical reasons for finding them very uncomfortable or difficult to wear. While masks are cheap, not everyone will be able to afford them, and they can run out. Masks make certain kinds of crimes easier to commit. Obviously, they impede reading facial expressions and, thus, communication. This issue is not a problem in contexts where one remains aloof, such as shopping, but could be more of a problem for people in a bar seeking to deepen their acquaintance. They also pose obvious practical problems for eating and drinking, and during the Pandemic, this usually meant that one was required to wear a mask while walking to a table or taking a seat at a bar, but then permitted to take it off, and likewise on aeroplanes. The cost–benefit analysis is also far from clear for teenagers and children in school settings, as we discuss below. Masks create environmental pressures and raise logistical issues around disposal. For short-term mask use, these are minor issues, especially if implemented during a period when other social distancing measures are limiting many of the aforementioned activities already. But assessing costs becomes more important if masking is maintained for a longer period, and at the extreme, some groups advocated permanent mask mandates in some settings.

However, perhaps the largest downside of masks was (ironically) unforeseen: that their visibility lends them to symbolism. In parts of the world, both policy and individual choice about masks became politically charged. The United

States led on this. Survey evidence from the United Kingdom indicates that compliance with public mask-wearing requirements, when these were in force, was high, often over 95% (ONS 2021), whereas "levels of adherence to test, trace and isolate" were "low" (Smith et al. 2021). Different explanations are possible but the visibility of masks is plausibly part of the story. In Asia, as we have already mentioned, masks excited no special comment, while in Africa they were greeted with a kind of amused tolerance, arising from the obviousness of more serious needs. For instance, it was sometimes possible to rent (not buy) a mask outside a store that required one for entry. In the wealthiest parts of the world, however, and in the wealthiest country in particular, wearing a mask came to imply a stance on many issues besides the effectiveness of doing so, while the political significance of mask policy outstripped either its public health impact or its human cost.

6.3 Masks Before Covid

The effects of mask-wearing on the transmission of any disease depend on how that disease is transmitted. This was subject to changing scientific opinion during the Pandemic, as was the effectiveness of mask-wearing. The use of respirators by medical professionals was already recommended in the context of influenza pandemic preparedness. A respirator filters air in both directions and protects the wearer as well as the environment, whereas a facemask is primarily designed to protect the environment from droplets and particles emitted from the wearer. Prior to 2020, the use of both masks and respirators was common in medical settings: surgeons wear masks to keep the environment sterile, and a carer intubating a patient with an infectious disease may need to wear a respirator.

However, there was no scientific consensus around population-wide mask use in response to an influenza pandemic. The *UK Influenza Pandemic Preparedness Strategy 2011* says it is "unresolved . . . how effective surgical masks or respirators might be in reducing transmission" (DH Pandemic Influenza Preparedness Team 2011, 37). It sees a place for both among medical professionals. However, it is sceptical about population-wide use of masks.

> Although there is a perception that the wearing of facemasks by the public in the community and household setting may be beneficial, there is in fact very little evidence of widespread benefit from their use in this setting. Facemasks must be worn correctly, changed frequently, removed properly, disposed of safely and used in combination with good respiratory, hand, and home hygiene behaviour in order for them to achieve the intended benefit. Research also shows that compliance with these recommended behaviours when wearing facemasks for prolonged periods reduces over time.
>
> *(DH Pandemic Influenza Preparedness Team 2011, 37)*

For this reason, the policy decision is as follows:

> In line with the scientific evidence, the Government will not stockpile face-masks for general use in the community.

This explains why the UK did not have stockpiles of masks, in common with other countries where mask-use was not already widespread (as it was in parts of Asia). The policy makes employers responsible for providing masks for employees where necessary, but provides no clues as to which kinds of businesses might face this necessity.

The policy is based on a systematic review, which makes for interesting reading from the present perspective. Several of the studies raise the possibility of recall bias, as they were conducted several months after the fact. This is an instance of the more general difficulty that mask studies rely on self-reporting, which is always a source of bias, in unpredictable ways. One study finds a positive effect in household use, as compared to use in educational settings. This, of course, is the exact opposite of Pandemic policies, and, except at Halloween, there cannot be many families that would wear masks around the home. One study finds what look like promising results in relation to SARS, which is a coronavirus, but it is far too small to draw any conclusions: three nurses with masks contracted SARS, while five without masks did not; and moreover, this is one of the studies vulnerable to recall bias. The evidence review only considers SARS because of the anticipated paucity of studies on influenza, but finds that the SARS research is no help. SARS has a very different epidemiology from influenza, the review explains, and moreover, the SARS studies it found were "poorly designed, had many weaknesses and were difficult to interpret"(Reza et al. 2012, 260–62).

In short, it is fair to say that masks were not seen as promising interventions prior to Covid-19. This makes it remarkable that masks have become a site of huge conviction for some. There are "pro-maskers," and there are "anti-maskers," and the strength of opinion is striking given that, just a few years ago (at time of writing), there were neither. The history of the evidence makes it unsurprising, on the other hand, that scientific opinion about masks is unsettled. Scientists do not change their minds en masse at the drop of a hat. Even the state of the scientific evidence has become a matter of dispute, and not everyone will accept that there even *is* a scientific debate to be had. That position, however, is demonstrably incorrect at the time of writing, and moreover, the process of scientific debate is what secures the conclusions of science and gives them special status. To put it another way, the strong stances taken on masks, either for or against, are not driven by a corresponding development of scientific evidence. Those who assert the opposite are wrong and are prematurely closing a debate that, if properly conducted, may hopefully yield the kind of consensus that is sometimes wrongly asserted to already exist. Prematurely closing a debate and ruling certain conclusions as off-limits is reminiscent of religious orthodoxy. This is the basis for our contention that there was a dogma concerning masks.

6.4 Wearer Protection Versus Source Prevention

Effectiveness can be understood in several ways. There is a distinction between *wearer protection*, in which a mask protects the wearer from becoming infected, and *source prevention*, in which a mask protects the wearer from being a source of infection for others. Both are implicated in controlling transmission, but the difference obviously matters for political purposes. Employees might refuse to work with unmasked employees if they regard the primary benefit of masks being source prevention. The use of masks in surgical contexts is driven by source prevention, and it is generally accepted (we believe) by pro-maskers that source prevention is indeed the primary benefit of facemasks, especially at the lower end of the spectrum of mask grade.

There is also a distinction between respirators and masks, with respirators being closely fitted and capable of filtering air both in and out, and thus protecting wearers. There are further distinctions within both categories depending on material, thickness, or ply count, and so forth. The best respirators are thought by pro-maskers to be effective in both source prevention and wearer protection. As we will see below, this is certainly true in a laboratory setting. The question is whether the evidence supports the use of masks or of respirators in public health interventions.

It is possible that the content of the Mask Dogma is true. A dogma is characterized, not by falsity, but by being inappropriately maintained in the face of recalcitrant evidence. Our focus here is on whether the maintenance of the assertion was appropriate, given the evidence. Our argument withstands the possibility that, in a decade or three, mask use has been proved conclusively effective in the prevention of Covid-19; and it withstands the opposite possibility too. In calling the Mask Dogma a scientific dogma, we are commenting on the attitude towards its propositional content. We are not commenting on its truth.

6.5 The Difficulty of Studying Masks

The evidence about masks and Covid-19 is mixed, and views about it differ, including among scientists. This much is common cause: that under laboratory conditions, masks are effective at source prevention, respirators are effective at both source prevention and wearer protection, and there is a clear hierarchy of performance from fitted higher-grade respirators at one end down to simple cloth masks at the other. Before the Pandemic, evidence for mask use among healthcare workers for influenza was not conclusive, while evidence for respirator use among healthcare workers showed effective wearer-protection. Pre-pandemic planning policy did not recommend face coverings outside healthcare settings, as mentioned above.

Since the arrival of Covid-19, many more studies of masks have been done, and considerable evidence has been produced. But the evidence on mask-wearing is complicated, because it varies along several dimensions, as follows.

- Whether mask-wearing or mask mandates are assessed. Good studies will be clear about what they are studying, while good reviews will usually consider the evidence for each separately, or employ some similarly clear approach.
- Type of mask used. Studies may differ in whether they distinguish between different types of masks and, if so, whether they seek to assess differences in effectiveness.
- Context of mask use. Studies may concern narrower contexts (e.g. aeroplanes) or broader contexts (e.g. general use by whole cities or even countries). The most significant distinction is between healthcare versus non-healthcare settings.
- Background of other NPIs. Masks are practically never used as the sole NPI. Before the pandemic, the main other NPI considered was hand-washing, but from 2020 on, there were many other NPIs in the picture. This background can bias individual studies and varies between studies.
- Wearer-protection and source-prevention. Obviously, there is some logical connection between the protection of wearers and the prevention of transmission at the population level, since transmission includes transmission to other mask-wearers. Nonetheless, in reality, the connection is complicated, and studies can provide information about one without providing information about the other.[1]
- Compliance. If compliance influences the effectiveness of mask use at the community level (as is the case for some other public health measures, such as vaccination), then low compliance will have misleading results for high-compliance contexts and high-compliance studies will have misleading results for low-compliance contexts. Also, compliance may vary greatly within a study.
- Reporting. It is hard to avoid relying on reported use, and this is intrinsically unreliable.
- Infection-verification. Studies may differ in the method used for verifying infection has occurred, using (reported) self-testing, clinical diagnosis, or laboratory-confirmed diagnosis. Laboratory confirmation is obviously desirable.
- Pre-intervention biases. Individuals who wear masks could be more vulnerable. Conversely, people disposed to wear masks might be disposed to take other precautions too. Alternatively, mask-wearers might feel invincible (as was feared at one time) and venture out more; and so forth. These all give rise to potential for bias (notably selection bias and collider bias[2]).

It is therefore not surprising that different people, including reputable scientists, take different views of the evidence. For instance, an update of an ongoing review of evidence on mask-wearing for respiratory diseases found that:

> Wearing masks in the community probably makes little or no difference to the outcome of laboratory-confirmed influenza/SARS-CoV-2 compared to not wearing masks. . . . Harms were rarely measured and poorly reported. . . . There is uncertainty about the effects of face masks. The low to moderate certainty of evidence means our confidence in the effect estimate is limited, and that the true effect may be different from the observed estimate of the effect. The pooled results of RCTs did not show a clear reduction in respiratory viral infection with the use of medical/surgical masks.
>
> *(Jefferson et al. 2023, 2)*

And a systematic review of mask mandates in children reached the following conclusion:

> Real-world effectiveness of child mask mandates against SARS-CoV-2 transmission or infection has not been demonstrated with high-quality evidence. The current body of scientific data does not support masking children for protection against COVID-19.
>
> *(Sandlund et al. 2023, 1)*

On the other side, one finds claims such as the following:

> Despite the [risk of bias], and allowing for uncertain and variable efficacy, we conclude that wearing masks, wearing higher quality masks (respirators), and mask mandates generally reduced SARS-CoV-2 transmission in these study populations.
>
> *(Boulos et al. 2023, 2)*

And:

> High-quality studies have shown that use of face masks in the community is associated with reduced transmission of SARS-CoV-2 and is likely to be an important component of an effective response to a future respiratory threat.
>
> *(Cash-Goldwasser et al. 2023, 1)*

What gives rise to these differences of interpretation?

6.6 Different Study Designs Yield Different Results

The explanation arises directly from the list of challenges facing mask studies. There are many kinds of epidemiological studies, and some are better than others at ironing these things out. The ones that do the best job tend to come up with rather small or no effect for masks. These are randomized controlled trials (RCTs, explained shortly). However, the ones that are more prone to bias come up with larger effect sizes, usually positive. These are observational trials, of which there are a great many kinds (explained shortly). Some scientists regard RCTs as clearly superior to observational trials and therefore conclude that the positive results in observational trials arise from some of the many biases previously listed. However, other scientists regard RCTs as too strict, because their strict standards can only be met for a small range of interventions (notably pharmaceutical ones, for which they are primarily designed). Some interventions just cannot be tested with an RCT. But that does not mean they do not work. Scientists inclined to this view see the evidence as showing, not that masks do not work, but that RCTs do not work for detecting the effect of masks. All the evidence shows, they believe, is that masks are among the interventions that RCTs just cannot test.

In a classic RCT, a large number of individuals will be randomly assigned to two groups (or more, if there are multiple treatments to be assessed). The groups themselves are then randomly assigned to receive the treatment being tested or no treatment. Neither the members of the group nor the investigators know which group is which (this is blinding). Thus, a placebo must be used for the non-treatment group. RCTs' designs must be pre-registered before the trial is conducted, so that there is no shifting of the goalposts to accommodate unexpected effects that could arise from some form of bias. There are many further details depending on the exact type of study, but in a nutshell, these measures are all designed to ensure that there is no way that an association between treatment and measured outcome can arise except by a causal effect of the treatment or an incredible coincidence.[3]

Many interventions are intrinsically ill-suited to such trials. No placebo can give one the impression that one is running an hour a day, for example. No ethical trial can randomly assign people to a treatment group that requires them to smoke (or, according to standard bioethical principles, to give up smoking). Some interventions are so effective that even if the study starts off blinded, it ends up not being so: consider a trial on the effectiveness of parachutes. Some causal factors are extremely hard or even impossible to intervene upon, such as race, sex, income, social class, occupation, and other biological or social characteristics.

In short, RCTs have significant limits on their range of applicability. Masks might be outside this range. There is no placebo for mask use. There are great difficulties implementing mask interventions in a randomized way, since individuals not compelled by law will usually choose for themselves whether or not they wish to wear a mask, based on their own opinion, rather than obey an investigator. The same goes for regions: no policy-maker is going to permit their jurisdiction to be randomly assigned to mask or no-mask groups for the purpose of a trial.

For this reason, RCTs on mask use fall short of the ideal standard. Here, for example, is the intervention specified for a well-known RCT conducted in Denmark:

> Intervention: Encouragement to follow social distancing measures for coronavirus disease 2019, plus either no mask recommendation or a recommendation to wear a mask when outside the home among other persons together with a supply of 50 surgical masks and instructions for proper use
>
> *(Bundgaard et al. 2021)*

This intervention was designed to test "whether recommending surgical mask use when outside the home reduces wearers' risk for SARS-CoV-2 infection in a setting where masks were uncommon and not among recommended public health measures." It is perfectly reasonable to point out: that the intervention might well result in low compliance; that other policies such as a strictly enforced mandate might result in much higher compliance; that the lack of general mask-wearing might result in social pressures to this effect; that the result therefore has no bearing on that policy; that the primary benefit of surgical masks is generally thought to be source prevention rather than wearer protection; and that social distancing measures might in any case reduce the observed effect of masks by reducing exposure. This illustrates a common problem with RCTs: in tightly specifying an intervention, the distance grows between that intervention and any real-world intervention. The investigators obviously could not change the law, so they did what they could: they made a recommendation. And this, obviously, is not what pro-maskers advocate; in fact, it is almost the opposite.

Observational trials, on the other hand, look for correlations between the treatment and the outcome of interest, and use statistical analysis to attempt to control for differences other than the treatment, leaving the treatment as the only plausible explanation of remaining differences in outcome, if any. These analyses *can* get things right, but there always remains a risk of residual confounding: that other differences between the groups, besides the intervention, account for the observed differences in outcome. Nothing guarantees that you have thought of them all, and thus nothing guarantees that you have eliminated them all in your analysis. It is obviously possible to conduct all kinds of comparisons between people who do and do not wear masks, between regions where masks are or are not mandated, between different types of mask use, and across different waves of an epidemic where masks were or were not used. However, other differences are bound to be present in any such comparison. Statistical ingenuity might eliminate them: we might, for example, be able to compare younger and older populations if we stratify by age. But we might not, and we might not know that we have not, since we might simply not know about some confounder that explains the observed difference in outcome.

Observational studies become more convincing the more of them are done, since it becomes less and less likely that the same confounders or biases operate across all of them. However, there is no definitive way of determining when enough evidence is enough.

Systematic reviews of the evidence are the most common way to establish what the evidence as a whole says. A systematic review needs to make a decision about what studies to include and what not to include, based on how good the study is. This decision is expressed in the *inclusion criteria*. The study may confine itself to randomized trials, or it may admit some observational trials, to a greater or lesser degree. Thus, there is no way to settle the issue by looking at the evidence as a whole. Reviews with stricter inclusion criteria include only RCTs and find little or no evidence of the effectiveness of masks, while reviews of observational evidence are more likely to find positive evidence.

Thus, a debate about the effectiveness of masks becomes a debate about epidemiological methodology, and debates about methodology cannot be settled directly by evidence. It is against this background that dogmatic insistence on a certain view of the evidence has arisen. We turn now to identifying the signs and symptoms of the Mask Dogma.

6.7 The Mask Dogma

We identify the Mask Dogma as follows.

Mask Dogma

> Masks are necessary and effective for responding to an epidemic wave in the Pandemic.

The sense of "necessary" in the Mask Dogma is simpler than the Lockdown Dogma. The point about masks is simply that they should be worn, not that doing so will keep a wave within manageable levels with respect to a healthcare system, and there is no particular goal (such as keeping the R_t below 1). The necessity derives from the general necessity to take measures against Covid-19. The rationale is that masks are effective, significantly so. Moreover, they represent great value, with few harms and substantial benefits. This is a contrast to the rationale for lockdown, which does not concern harms and benefits, but focuses overridingly on hitting a certain threshold.

6.8 Bad Causal Inferences

Whenever nuance is necessary, simplification is possible. Over-simple causal inferences about masks were disappointingly common during the Pandemic among advocacy groups, commentators, and even scientists who ought to have known better. The most common arose from the temptation to compare countries and attribute differences in Covid-19 outcomes to differences in mask-wearing, despite the obvious presence of many other differences. Comparisons between countries can be informative, but they require careful interpretation. They are best conducted with reference to eventual endpoints, since – as we discussed in the previous chapter – a country can appear to be "doing well" at one point and then end up "doing badly," and vice versa. And their ability to discriminate between different causal factors, let alone between that subset of causal factors comprising interventions, is inevitably limited. Many of the same points apply to the inter-country comparison of lockdowns, although that comparison has an important role to play in assessing whether lockdowns were universally necessary to achieve a given outcome.

The debate about masks, on the other hand, was simply about whether they had a significant, or any, effect, and inter-country causal inferences made in this context were very crude. The advocacy group "Masks4All" produced a meme showing how countries in Asia that enforced masking enjoyed much better outcomes than countries that did not (see Figure 6.1).

Although the approach is obviously simplistic, this is part of its rhetorical power. The crude markup looks like it was done using a mobile phone image editing function. That is all you need, says the meme, to see the effect of masks. It is just obvious. And many of the millions who shared the meme agreed.

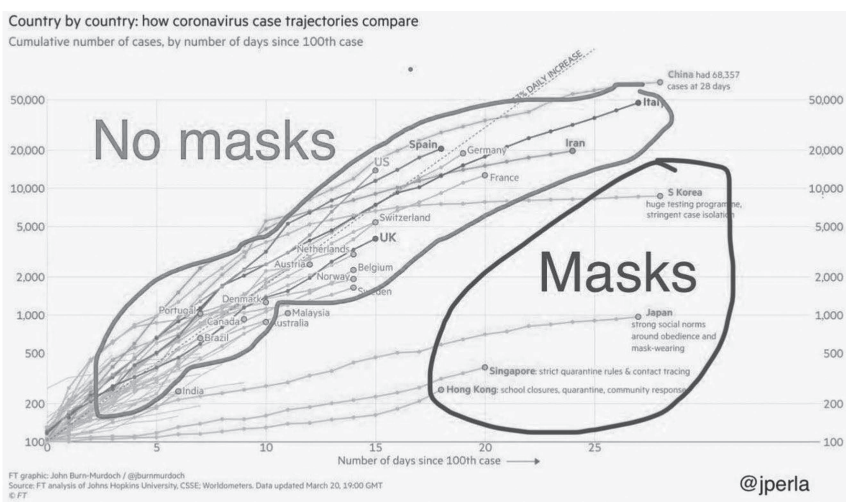


FIGURE 6.1 Meme propagated by advocacy group Masks4All

Of course, the authors set themselves up for lampoon (see Figure 6.2), because the masking countries shared many other properties, including some that were obviously not causal. Simple comparisons of this kind are not altogether useless for causal inference, but without context, they mean nothing.

What is more concerning is that similar inferences were made by scientists. For example, when an RCT on masking in Denmark was published, a science communication NGO published eleven expert opinions (Science Media Centre 2020). There were many reasons to be critical of that study, or at least of a certain interpretation of that study, and many of the opinions identified these and were reasonable to do so. However, the leading opinion is by an "expert" with a background in virology, rather than epidemiology, and his answer is essentially a sophisticated version of the above reasoning:

> During the ongoing COVID-19 pandemic, the compliance with universal masking in these Southeast/East Asian countries continues to be very high relative to Western countries – with clear benefits for virus control.
>
> *(Science Media Centre 2020 – see Julian Tang's comments)*

The inference has a more sophisticated appearance than the above meme because the gist of the answer is that compliance matters for the effectiveness of masking, and compliance was low in the Danish study. However, this certainly does not warrant the claim that the effect of masking *would* have been higher if the study were done in East Asia due to higher compliance. And it certainly does not warrant the claim cited above, which is more general still. Several commentators

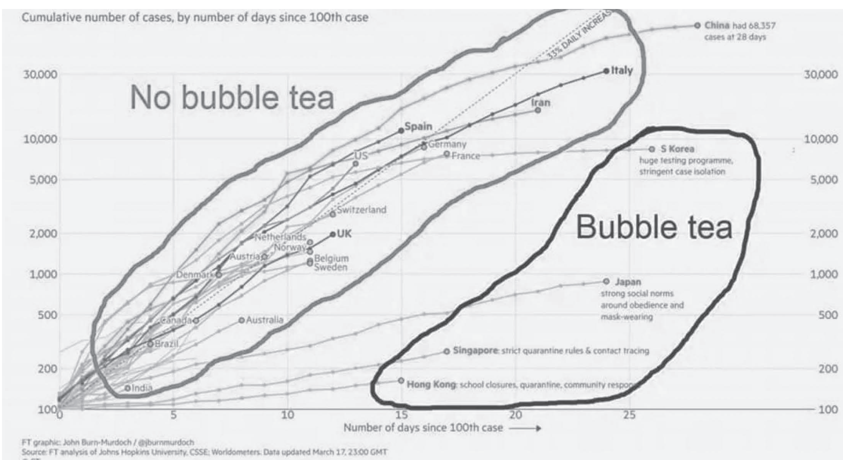


FIGURE 6.2 Meme satirising one propagated by the advocacy group Masks4All

mention compliance (as do the study authors), but to say that higher compliance would have resulted in a larger effect is merely to assume that masking is effective, which is to assume the truth of the hypothesis being tested. That is not science, and especially not when – as we have already indicated, and as many of the remaining (and more credible) experts indicate – there are plenty of important points to make about the Danish trial.

Many scientists made such errors, and, to the extent that they arise from cognitive biases, we can generally expect that scientists are no less prone to do so than the rest of us when they are reacting swiftly (as they might in an interview, or when writing an op-ed in a hurry) rather than thinking carefully (as they might in their own research) (Kahneman 2011). There are many "ecological" studies of differences between countries, and some of these either imply causal conclusions or are obviously motivated by seeking them. These studies are rightly excluded in good reviews, on exactly the basis just given (Boulos et al. 2023, 4).

Lest the reader think we are being too dismissive of these inferences, here is another example of the dangers of drawing inferences from inter-country comparison. Czechia implemented mask mandates during periods when most of Europe did not, and initially enjoyed relatively good outcomes (see Figure 6.3).

It is reasonable enough for an association of this kind to spark a hypothesis. However, Czechia subsequently suffered much worse outcomes than many nations that did not mask (see Figure 6.4), illustrating that such evidence ought not to be treated as conclusive.

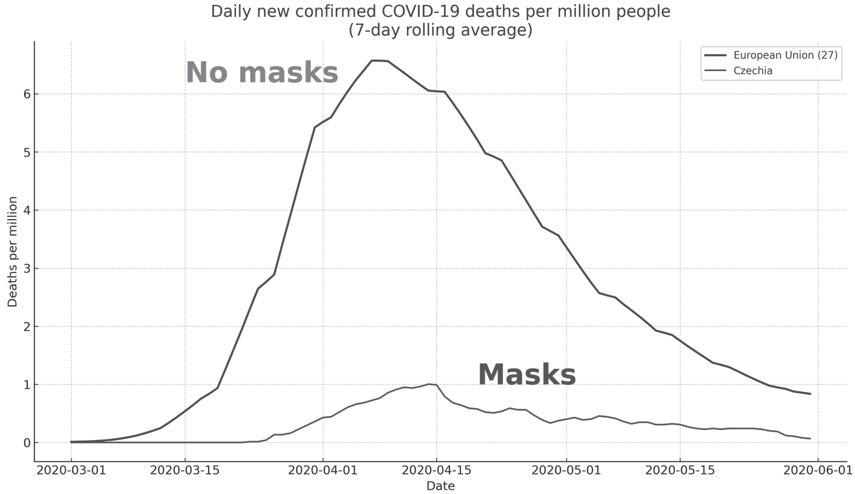


FIGURE 6.3 Favourable comparison between Czechia and the EU during the first wave

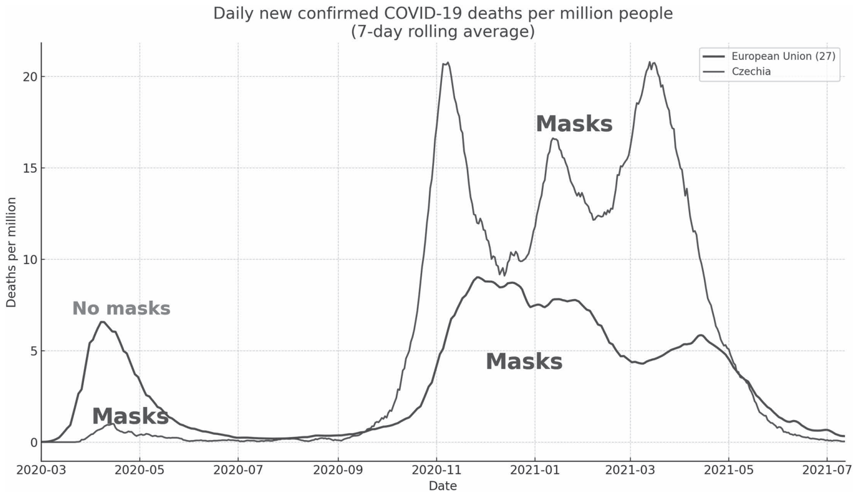


FIGURE 6.4 Unfavourable comparison between Czechia and the EU in the second wave

The evidence from the first wave did not clearly tell for or against mask-use, but it was interpreted that way by some. The second wave showed that this was an error. It is reasonable enough to treat the first wave experience as a *prompt for investigation*, especially in a context where solutions are urgently sought. However, associations of this kind ought never to have tempted us to make positive causal inferences. The fact that such causal inferences were as popular as they were indicates that certain propositions were insensitive to evidence by being insensitive to its significance or, to use a more popular term, its quality. A good scientist ought not to say that masking had "clear benefits for virus control" in Asian countries, because the matter is clearly not clear. Many scientists made no such mistake, but some did; and this we identify with adherence to a dogma.

The more sophisticated version of the mistake is to reject restrictions on "quality" of evidence on an ad hoc basis, and we will now turn to this discussion.

6.9 The Dilemma of Disappointing Evidence

As it has turned out, the evidence on masks presents many scientists with a dilemma. Evidence is commonly ranked. The "best" evidence does not support the view that masks have much or any effect on Covid-19 transmission. "Lower quality" evidence does, however. Should standards of evidence be "lowered" so that this "lower quality" evidence plays a significant weight in determining masking policy in future pandemics? This can be achieved by attacking RCTs in various ways, either individually or *en masse* (both have been done in the mask context), or simply by not applying tests for bias that would exclude observational studies. But the former smacks of *ad hominem*, since if the RCTs favoured mask-use, proponents would doubtless be shouting the fact from the rooftops, while the latter is hardly a good look for scientific rigour. So, should the RCTs and the systematic reviews that select only them be allowed to determine future policy? That, too, is not a good look, since it implies ignoring the vast majority of research on mask-wearing. It also does no credit to the solid evidence on the physical effectiveness of masks in laboratory settings.

The dilemma of disappointing evidence is not unique to this context. In fact, it is common. Experiments often "go wrong," and need to be done again. It is common for investigators to refer to "failed RCTs" when the expense and trouble of running a trial show that the intervention being tested does not have the effect whose possible existence prompted the trial. Especially in biomedicine, research usually has a hoped-for outcome.

When the hoped-for outcome does not materialize, however, one can respond in one or more of four ways, broadly speaking:

1 reject the hypothesis under test (e.g. that masks are effective),
2 reject some other "auxiliary hypothesis" (e.g. that recommendations for mask-wearing are likely to produce behavioural change),
3 reject the integrity of the trial, study, or experiment that produced the result (e.g. by alleging fraud, incompetence, etc.), or
4 reject the method employed in that trial, study, or experiment.

It is not always easy to draw the line between a hypothesis being disconfirmed and a hypothesis merely failing to be confirmed. No hypothesis faces experience on its own, but as part of a larger web, including other theoretical components as well as mundane ones (concerning whether the instrumentation is working, whether the lab assistant is lying, etc.). Faced with an unexpected result, there are always options, and rejecting the hypothesis under evaluation is just one of these options. And it is often the wrong one. Scientific findings are often anticipated long before they are confirmed, and it takes persistent and ingenious experimentation to confirm a hypothesis that is suspected to be true. If scientists rejected hypotheses too readily in the face of recalcitrant evidence, then many scientific discoveries would never have been made.

We have already seen how the results of different types of studies about the effectiveness of mask-wearing can lead to a debate about the relative merits of those study designs. If one accepts "low quality evidence," then one risks incurring unnecessary costs, and if one rejects it, one risks rejecting a potentially life-saving intervention. The question in the mask debate is what to do about this dilemma. We once again see evidence of dogma in the way some scientists have responded to this question.

6.10 Precaution

A convincing response to this kind of uncertainty is to err on the side of caution. Masks are relatively low-cost (including in the broad sense that includes all kinds of harm), and the potential benefit is large. Therefore, until more is known, they should be used.

A precautionary argument replaces risk–benefit analysis in circumstances of uncertainty. It is sometimes formulated as the application of a *Precautionary Principle* (PP), which is the subject of considerable philosophical debate, but has an obvious intuitive pull. You are not sure if a car is coming or not, so you look twice before you cross. You do know enough about the likely costs to you of being hit by a car to conduct a cost–benefit analysis, but you know it is far greater than the cost of looking twice, so you do not need to. If we are unsure, but it is clear that a far greater risk attends one kind of error than another, then we should err on the side of caution.

This argument was mounted early in the Pandemic at a time when early exploratory studies on masks showed reasonably promising results, while the prevalent public health advice (including the WHO's position) maintained that mask use was unnecessary or even counter-productive. At that time, it was a good argument: one could dispute the claims about the harms of masks, but the evidence in regard of potential harms was certainly less at that time than the evidence of potential benefit, and there was clearly at least a cogent precautionary case for mask use. However, precautionary arguments are time-dependent. It is not reasonable to run a precautionary argument three or four years later. At exactly what stage the precautionary argument became unreasonable is a matter for variation of opinion, but our point is simply that as the evidence comes in, the claim that we need to act cautiously in the face of uncertainty dwindles. That is so even if we remain uncertain. That is because, whatever the effect of masks, the plausible *size* of that effect shrank during the Pandemic, along with the plausible risk. Cost–benefit analysis replaces precautionary reasoning at some stage as the costs become smaller.

The plausible effect size shrank for two reasons. One is just obviousness. Places that did not implement mask mandates did not exhibit horrendous outcomes of a truly obvious kind. Thus, the idea that there was a major risk to not wearing a mask loses traction. Of course, any avoidable loss of life is still a major human cost. The question is whether this cost is outweighed by the benefit. We can no longer say that we have no idea what the cost is, however, since in broad terms we know it is not enormous.

Second, the evidence for the effectiveness of masks has indicated a smaller effect size in each subsequent year since 2020. Evidence can, of course, become stronger while finding a smaller effect size, so this is not the same as saying that the evidence has weakened. Rather, it means that as the certainty around masks has increased due to an accumulation of evidence, it has pointed to smaller and smaller benefits. It is an interesting question how to interpret this finding, and indeed, it is not a finding we have previously seen discussed at all. Perhaps it is illusory, or will reverse. Even if so, it undercuts the use of a precautionary argument and instead replaces it with a cost–benefit argument.

6.11 Any Costs Are Worth It?

One might still insist that *any* plausible costs of mask-wearing are worth the benefit, since the benefit concerns saving lives, whereas the costs concern matters of convenience. This is a poor argument for two reasons. First, it is completely insensitive to evidence about the costs of mask-wearing. Humans use facial expressions to communicate. Children learn and experience love through facial expressions. Crimes can be committed with masks on. The evidence on harms is growing, too, and it cannot be asserted that this evidence will never amount to much.

Second, trade-offs between life and convenience are ubiquitous. We could set speed limits at 5km/h, and perhaps there are people who would advocate for this. But most of us would prefer a speed limit that guarantees some level of road death, on the basis that it is more convenient. To trade off life and convenience is not abhorrent, but normal.

This does not show that the risks of not wearing a mask are small enough, and the costs are high enough, to justify not wearing a mask. But it shows that the calculation must be made. One cannot simply dismiss the entire calculation when one side involves mortality. As a famous phrase has it, the essence of strategy is deciding what not to do.

6.12 Insisting That High-Quality Evidence Exists

Another reason to think there is a Mask Dogma is the insistence in some quarters that high-quality evidence to support mask use in fact does exist.

For example, a recent (at the time of writing) and prominently published article makes the following claims.

i) “Well-designed observational studies have demonstrated the association of mask use with reduced transmission of SARS-CoV-2 in community settings, and rigorous evaluations of mask mandates have found substantial protection” (Cash-Goldwasser et al. 2023, 1).
ii) “Literature review revealed many high-quality observational studies demonstrating the association of face mask use in the community and of mask mandates with reduced spread of SARS-CoV−2” (Cash-Goldwasser et al. 2023, 1).

iii) "Robust available data support the use of face masks in community settings to reduce transmission of SARS-CoV-2 and should inform future responses to epidemics and pandemics caused by respiratory viruses" (Cash-Goldwasser et al. 2023, 1).
iv) "High-quality studies have shown that use of face masks in the community is associated with reduced transmission of SARS-CoV-2 and is likely to be an important component of an effective response to a future respiratory threat" (Cash-Goldwasser et al. 2023, 1).

The first claim states the *existence* of good studies; the second states that they are *numerous*; the third argues that they *should inform future responses*; and the fourth implies that masks *ought to be used* in future responses.

The third and fourth claims are loose, and we should not be detained by them. Saying that evidence should "inform" future responses is vague, and saying that associations should ground interventions is an unfortunate error. Neither mentions costs at all. These statements do not help with an assessment of the evidence. They are attempts to influence policy: to exert epistemic power on the basis of implied IPA.

The first two claims are more interesting. If, as they say, well-designed observational studies had indeed demonstrated associations between mask use and lower transmission, and between mask mandates and "protection," then this would imply that a number of other reviews are wrong. As already indicated, there exist systematic reviews concluding that *no* good studies (or at least no *very* good studies) support mask policies. (This includes the more sober reviews concluding in favour of mask use, as we shall see below.)

Unfortunately, the review does not make a convincing case that the claims are true. In fact, it would not be entirely unreasonable to suggest that this review itself is of low quality. It has no inclusion criteria and indeed no clear methodology. It identifies and builds a narrative around a number of studies showing positive effects. Not a single study showing a negative or neutral effect is cited, even though they exist, and because there are no inclusion criteria stated, there is nothing to gainsay an inference that they were excluded because they did not support the conclusion of the review. Moreover, about half the citations are from 2020, between two and three years prior to this paper's 2023 publication. Yet the number of mask studies has increased. What has decreased is the effect size reported in those studies (as we will explain below). Again, this is uncomfortably compatible with the possibility of cherry-picking.

The review never explains what makes the studies it surveys "high quality." Earlier studies, on which the review substantially relies, were usually of lower quality for the very good reason that they tended to be exploratory, convenience-based, and designed to prompt further research rather than settle a question. The review rejects the idea of a meta-analysis on the basis that studies are too heterogeneous, which is reasonable enough. But it does not replace this with any systematic assessment whatsoever of the quality of evidence. One might reject the idea that studies must seek to approximate RCTs, but one needs *some* criteria for evaluating a study before relying upon it. Nothing in this review prevents the inclusion of a horoscope reading.

When one digs into the studies cited, they do indeed turn out to be of variable quality, and inferences drawn from them are not always appropriate. For example, consider this paragraph.

> Observational studies have demonstrated the effectiveness of mask use to prevent transmission of SARS-CoV-2 on airplanes,[27] in schools,[28,29] and among household[30] and community[17,18,31] contacts of individuals with COVID-19. A COVID-19 outbreak on the USS Theodore Roosevelt aircraft carrier was particularly instructive. Ships are high-risk environments for respiratory disease outbreaks because they bring people together for prolonged periods in often poorly ventilated close quarters. The outbreak on the USS Theodore Roosevelt occurred early in the COVID-19 pandemic, before crew members would have had immunity to SARS-CoV−2. More than 80% of those who reported not masking were infected; the odds of infection were 30% lower among those on the ship who reported masking.[32]
>
> *(Cash-Goldwasser et al. 2023)*

This is a compelling paragraph, if what it says is true. However, much of it is misleading.

To begin with, the phrase "observational studies have demonstrated the effectiveness of mask use" is far too strong. All the underlying reports describe significant limitations, even those that (perhaps inappropriately) include clear recommendations about mask policies.[4] For example, the conclusion of the paper on *USS Theodore Roosevelt* states, "Use of face coverings and other preventive measures could prevent transmission." This was an opportunistic[5] study done at short notice in 2020, with a view to generating information rather than testing hypotheses; it does not pretend to demonstrate anything. This paper identifies several of its own limitations, as do the others mentioned in the passage. Several are case–control studies, which have an important place, but which are subject to well-studied biases. They might form *part of a strong case*, but this is not the same as saying that they have *demonstrated* anything.

Likewise, the referenced study of aeroplanes does not regard itself as demonstrating anything. It includes cases where strict mask policies were in place and "an enormous number" of Covid-19 cases occurred. The study concludes "At present, based on circumstantial data, strict use of masks appears to be protective," and strongly urges further research. Again, the remarks about the *USS Theodore Roosevelt* are highly unsatisfactory, since – as the study itself points out – there could be more than one explanation for the differential odds of infection among those on the ship who reported being masked. The paper identifies four limitations and concludes that "Use of face coverings and other preventive measures could prevent transmission." This is not the same as saying that effectiveness was demonstrated.

Similar exercises can be conducted with the other papers cited. They do not evince anything like the confidence of the passage that cites them. This is emphatically not to say that they are "bad science," or "low quality" in any pejorative sense: they are *uncertain* science, and generally they are clear about this. Cross-sectional and case-control study designs are well-represented in the group, and these designs are not used for strictly testing or "demonstrating" hypotheses, but for generating them, and building a case which would then be subject to more extensive – and expensive – types of study.

Let us summarize the issues with this review. First, a full 20 of the 47 cited papers, including the two just discussed, are studies done in 2020. The paper was published on 31 October 2023, and more mask studies were published in 2021 and 2022 than in 2020. This implies a selection whose rationale is not stated. Second, as we will mention later, the average effect size seems to have become smaller in successive years: 2020 studies show the largest effects. Third, no studies with contrary findings are cited, even though they exist, and are mentioned in other (better) reviews. Finally, there are no inclusion or exclusion criteria stated. There is nothing to gainsay an accusation of cherry-picking. The paper simply provides a list of studies supporting the paper's conclusions. It is unacceptable to draw any conclusions in the name of science from such a methodology.

6.13 Unclear Reasoning

There are positive assessments of mask use that avoid all the above problems. For example, the Royal Society's report on NPIs is informed by an excellent systematic review of observational evidence on mask use, which is clear-sighted about the problems with observational evidence. It concludes in favour of mask-use regardless.

> Study designs commonly suffered from a critical risk of bias. The effects measured in each study were variable in magnitude and generally of low precision. Nevertheless, the weight of evidence from all studies suggests that wearing masks, wearing higher quality masks (respirators), and mask mandates generally reduced the transmission of SARS-CoV-2 infection.
>
> *(The Royal Society 2023, 29)*

The underlying study finds that 91% of observational studies were at "critical" risk of bias (ROB) in terms of the ROBINS-I system for assessing the quality of observational evidence (Boulos et al. 2023), about which more below. Critical ROB exists when "The study is too problematic in this domain to provide any useful evidence on the effects of intervention." The review makes appropriate distinctions between source-prevention and wearer-protection, and between healthcare and community settings, and provides forest plots of the results, including the (few) studies that show no effect. In short, it does what a review of the evidence ought to do. Even if one debates the inferences drawn, one does so on the basis of clearly presented evidence.

Yet even here, we see evidence of dogma. The Royal Society's report is obviously designed to inform policy, and its recommendation is not clearly connected with the evidence. If the evidence is unclear, then why is the recommendation in favour of mask use? There is a gap. There is also a correlation between the size of the effect found and the year in which it was found, which the analysis missed.

6.14 Ignoring the Real Possibility of Bias

Striking in all the positive reviews of evidence of mask use is the absence of serious discussion of the possibility that the positive observed effects are, in truth, a result of bias. Yet there are reasons to think there might be. Consider this hypothesis: the lower the quality of the study, the more likely it is to find a positive effect of masks and the larger the effect. Has this hypothesis been tested? We do not think so, and it is worth testing. There is a tendency for earlier studies to observe larger effect sizes, and earlier studies tended to be lower "quality," although often for perfectly respectable reasons, as we have already explained.

What is more, the same point even seems to hold for RCTs. A well-publicized RCT of masks in Bangladesh reported a significant protective effect (Abaluck et al. 2022). When this was reviewed for Cochrane, the authors found a high risk of bias in five of the six categories considered and therefore excluded it. These are not mere insults: the authors supply their reasons, and they are good reasons: reasons to suspect selection bias; lack of blinding; incomplete data; and selective reporting. However, they overlap with problems that mask supporters raise with RCTs finding *no* effect, and the Cochrane report rejects those studies for the same reasons.

This is another indication of an inappropriate attitude to recalcitrant evidence. One cannot dismiss the studies one dislikes on the basis of flaws that are shared by the studies one likes, and whatever else one says about it, the Cochrane approach is consistent. This point tells against pro-maskers who rail against over-reliance on RCTs, because some RCTs *do* find an effect of mask-wearing, but they tend not to make the grade for inclusion in systematic reviews. A possible inference is that the finding of a positive effect is correlated, not with study design, but with risk of bias itself – a damning inference, if supportable.

The aforementioned Bangladesh mask study did not make the grade for Cochrane, but it was eventually published in *Science*, despite widely aired and substantive methodological criticisms that predated peer review. The actual findings – a roughly 11% relative reduction in symptomatic seroprevalence overall, and somewhat larger reductions (~35%) limited to older adults – were modest and of borderline statistical significance. Furthermore, the absolute reductions were notably small, prompting critical commentary on statistics blogs and elsewhere regarding the selective framing and presentation of these results. In a blog post, statistician Ben Recht discussed the omission of raw seropositive case numbers and the potential overstatement of the intervention's effectiveness (Recht 2021). The absolute numbers of cases prevented were not initially disclosed, thereby probably exaggerating the public perception of the intervention's effectiveness when the findings were first reported.

This particular issue was corrected, but there were others, and the paper was still published. Subsequent reanalysis identified significant sampling biases in the trial, suggesting that staff behaviour led to imbalances between treatment and control groups, undermining causal inferences about mask efficacy (Chikina, Pegden, and Recht 2022). Another study analysed mobility data during the Pandemic in Bangladesh and found evidence of risk compensation behaviour following mask mandates, which could have influenced the outcomes of mask efficacy studies (Wadud, Rahman, and Enam 2022). Arguably, the study's scale, topicality, and alignment with prevailing Pandemic-era dogma regarding masking outweighed the methodological and interpretive issues raised by statisticians and critics. These critiques were publicly accessible and widely shared. The fact that the study was published in *Science* notwithstanding, and continued to be heavily publicized as showing that masks save lives, is evidence of a mask dogma, and of a broader epistemic environment characterized by scientific orthodoxy.

6.15 Conclusion

We have argued that the Empirical Content of the Mask Dogma was maintained in the face of recalcitrant evidence. We have not shown it is false. But we have shown that it was endorsed too strongly, wrongly foreclosing debate. Those who did, or still do, endorse it strongly must reject evidence arising from RCTs and endorse evidence that, on any clear-sighted evaluation, is at risk of bias. Sometimes, they endorse terrible causal inferences. Precautionary stances are outdated, and outright rejections of cost–benefit analysis are naïve. We looked at a low-quality review of the evidence published in a prominent journal, which made no mention at all of contrary evidence and ignored all the caveats in the evidence that is cited. We discussed a much better review that took a clear-sighted view of the evidence, but this review reached a positive conclusion without making clear the chain of reasoning connecting the evidence to the conclusion. We have not been able to find any work that properly addresses the possibility that there is a real connection between risk of bias and probability of finding a positive result, or the size of this result. It is a real possibility that various other factors – predispositions to take more care, for example – account for the effect found by observational studies. This is a possibility that the scientists most strongly favouring mask use appear not to take seriously. The possibility exists, too, that RCTs are actually better than observational studies at reflecting how masks are used "in the wild," or the real-life effect of mandates or recommendations. None of this shows that masks do not work. It shows, however, that the strength of opinion one encounters in relation to masks bears no relation to the strength of the scientific evidence.

Notes

1 For example, a trial of mask recommendation in 2020 made "no assessment of whether masks could decrease disease transmission from mask wearers to others" (Bundgaard et al. 2021).
2 Selection bias occurs when the outcome is already more or less probable in the intervention group even before the intervention (e.g. healthy people are more inclined to take up exercise, creating an association between exercise and health prior to any effect of exercise). Collider bias occurs when two independent variables are correlated with an intervention creating an inverse association between the two variables among the intervention group (e.g. if acting ability and attractiveness are independent causes of success as a Hollywood actor, then attractiveness will be associated with low acting ability and vice versa among Hollywood actors, since the chance of finding both properties together is lower than the chance of finding one – like rolling two sixes rather than just one).
3 There is no requirement that the *size* of the effect be large: it could be that a trial convincingly shows that a certain drug prevents a very small number of early deaths from a certain disease. One common criticism of pharmaceutical development is that it results in incremental treatments of this kind.
4 It is doubtful whether *any* study can demonstrate effectiveness without replication.
5 No criticism is implied. It was reasonable to take this opportunity.

7

SCIENTIFIC DOGMA AND COVID-19 VACCINES

7.1 Introduction

Vaccines have been controversial since their invention. They are administered to healthy people by medical professionals whose sworn priority is to do no harm, yet they do sometimes do harm, and not all healthy people believe the reassurances that the benefits are worth the risks. The Covid-19 vaccination programmes certainly saved a large number of lives. It has sometimes been objected that a correlation exists between a low vaccination status and poverty, which accounts for the (inverse) correlation between vaccine status and mortality. We show that this is not correct, because including both in a combined index improves the correlation. Nonetheless, exaggerated claims were made for the effectiveness of vaccines. To explain these, we posit the Covid-19 Vaccine Dogma: Covid-19 vaccines are safe, effective, and necessary. The meaning of all three terms is not always transparent. Not all scientific authorities subscribed to the Vaccine Dogma, and over time, its influence waned. Nonetheless, it existed, and still persists in some quarters, resisting recalcitrant evidence.

Vaccination against Covid-19 saved millions of lives, but it was also the subject of a scientific dogma. Vinay Prasad and Alyson Haslam argue that it represented a great success for science, but a great failure of implementation, particularly in America (Prasad and Haslam 2024). We have a similar stance, but would not express it as a distinction between science and implementation, because the two were strongly linked: erroneous implementation was often based on erroneous science. The development of the vaccines was indeed a remarkable achievement. However, the success of the science that developed the vaccines

DOI: 10.4324/9781003732495-8

was not matched by *some* of the science that influenced subsequent policy – that science we call the Pandemic Orthodoxy.

7.2 The Successes

There is no doubt that Covid-19 vaccines saved millions of lives. During the Delta wave in 2021, it was clear that countries with higher vaccination levels saw a significant drop in infection fatality rates (IFRs). This significant drop in IFR allowed the dropping of most suppression measures. Covid-19 was defanged. Vaccines played a major role in removing the fangs – where they were used.

There is a significant body of evidence to this effect. Compare South Africa, which had not vaccinated widely prior to the Delta wave, to the UK. In South Africa, CFRs (and hence also IFRs) remained largely unchanged during the Delta wave. While some ascribe the drop in IFRs in the UK to high attack rates, it must be noted that attack rates were even higher in South Africa (SA ~40% (SACMC 2021) versus UK ~25% prior to Delta (ONS 2023)), and yet there was no drop in IFRs (inferred by no drop in CFRs) in South Africa.

In South Africa, with a population of 59 million, excess deaths during the Delta wave amounted to 113,000 (about 0.19% of the population), while in the UK, with a population of 67 million, excess deaths during the Delta wave amounted to 27,000 (0.04% of the population). In both countries, approximately 25% of the population was infected during the Delta wave (SACMC 2021; ONS 2023). By the end of September 2022, the Delta wave made up 38% of all pandemic excess deaths in South Africa, but only 16% of all pandemic excess deaths in England (see Figure 7.1).

England had approximately 15,000 deaths with Covid-19 on the death certificate during the Delta wave (0.027% of the population). In a counterfactual scenario, where IFRs by age remained the same during the Delta wave, England would have had 74,000 Covid-19 deaths (0.132% of the population). This counterfactual scenario points to the vaccine having saved 59,000 deaths in England during the Delta wave (0.105% of the total population).

The reduction in IFR in England over time in each age band, coinciding with vaccination, is a diachronic analysis. The comparison between South Africa and England is a combination of a synchronic and diachronic analysis. This crude inter-country analysis on its own counts as weak evidence, but combined with the diachronic analysis from England, the evidence becomes significantly stronger.

It is sometimes objected that poverty status was inversely correlated with vaccination status: poorer people were less likely to be vaccinated, even in areas where vaccines were available. Poverty is also correlated with Covid-19 mortality. Does this explain the apparent effect of vaccines?

Sceptics sometimes claim that poverty explains *all* of the mortality differences, concluding that the vaccines had no beneficial effect. Vaccine supporters,

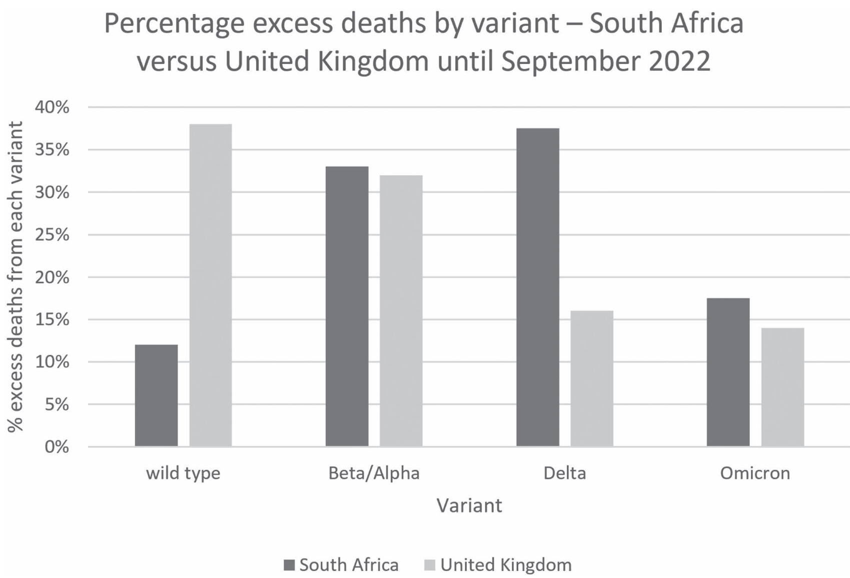


FIGURE 7.1 Proportion of pandemic excess deaths per variant, UK and South Africa

on the other hand, sometimes *completely ignore* the effect of poverty, thereby inflating the benefits of the vaccine. The truth is somewhere in the middle, where both low poverty levels and high vaccination levels were beneficial in equal measure.

One can see this by analysing the differences between all 50 US states and seeing that combining the factors improves the correlation. Data from the US show that lower levels of Covid-19 mortality were associated with lower poverty levels and higher vaccination levels (Bollyky et al. 2023). Our own analysis shows a correlation coefficient of 0.55 between poverty levels and age-standardized Covid-19 mortality (Figure 7.2) and 0.58 between vaccination levels and age-standardized Covid-19 mortality (Figure 7.3).

When combining the effects of poverty, vaccination, obesity, and stringency with a weighting of 6:6:2:1, the correlation coefficient increases to 0.76 (see Figure 7.4).

This combination displays the strongest correlation. If including a variable improves the correlation, then, while this is not by itself evidence for causation, it contradicts the hypothesis that an observed correlation between the other variables arises due to a correlation with the additional variable. The correlations between poverty status, vaccination status, and mortality do not undermine but rather support the claim that Covid-19 vaccines were effective.

It is clear that vaccination saved many lives during the Pandemic. However, it is also the case that the number was sometimes exaggerated. This is unfortunate

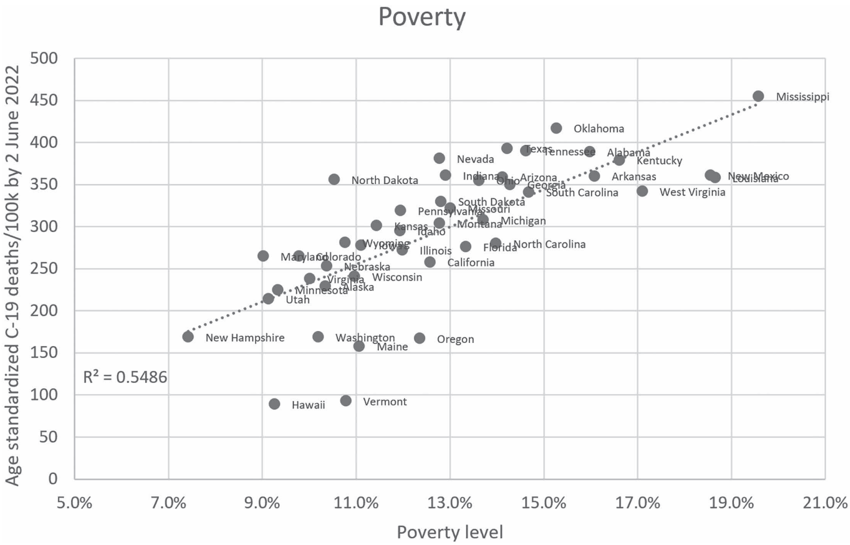


FIGURE 7.2 Poverty and cumulative mortality for US states

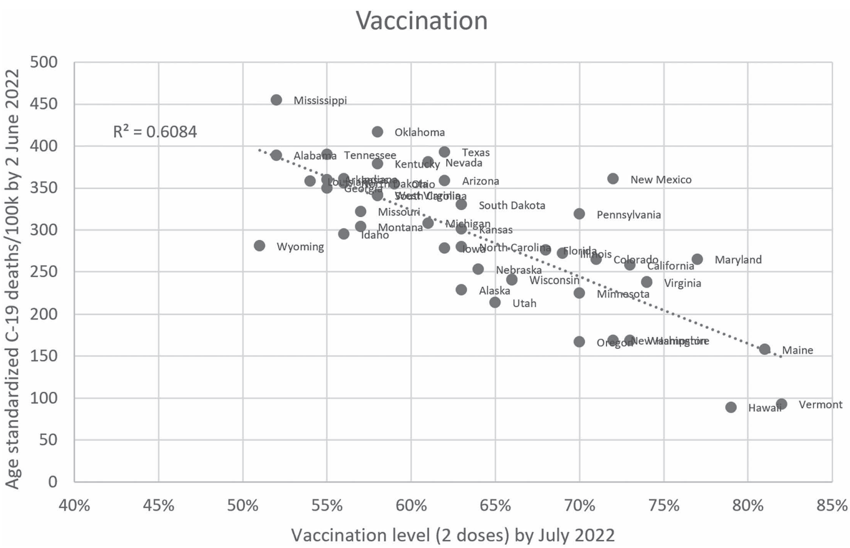


FIGURE 7.3 Vaccination status and mortality for US states

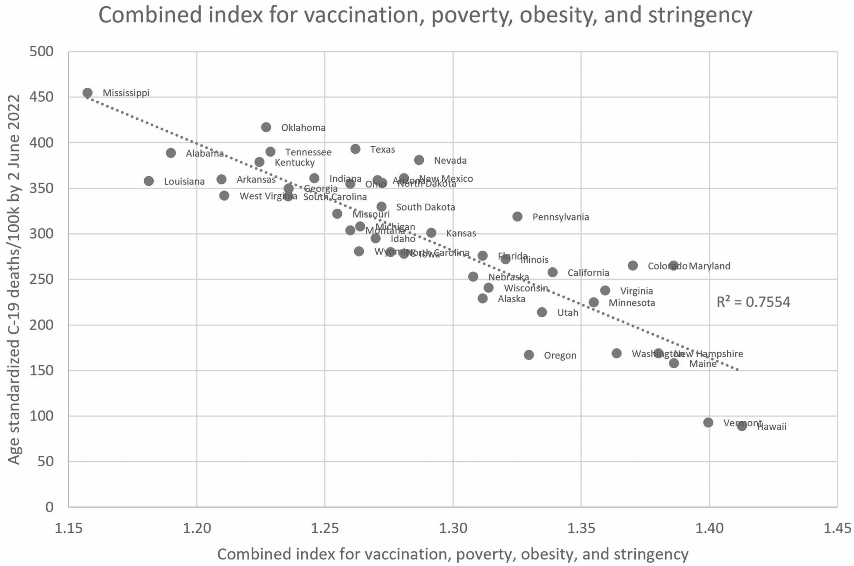


FIGURE 7.4 Combined predictive index for mortality in US states

because it is unnecessary. For example, there is a widely cited study that estimated that vaccinations averted 19.8 million deaths globally (0.25% of the global population) during 2021 (Watson et al. 2022). However, this study assumes a level of vaccine protection against infection, disease, and death obtained from other observational studies. This study can therefore not be used to validate those observational studies, as its own output is merely an extrapolation from those studies. This estimate is high in our opinion, considering that England most likely saved 0.105% of its population during 2021 by rolling out vaccinations prior to the Delta wave. Compared to the rest of the world, England was ahead in terms of vaccinations. England also relaxed most restrictions post-vaccination, allowing Delta to spread, and therefore had a relatively high attack rate of 25% during the Delta wave, compared to most other countries.

A saving of 0.1% of the population, in countries like England that vaccinated before Delta (or in countries with attack rates below 25% before vaccination), would still amount to millions of lives saved globally. To claim that the global savings were 2.5 times higher globally than in the UK is, therefore, as unnecessary from the perspective of validating or justifying vaccination as it is implausible from an evidential perspective.

Another area of exaggeration concerns the end of the Pandemic. Vaccines played an important part in bringing the Pandemic of serious disease to a close, but they are not the whole story. Many people had never received any Covid-19 vaccine by the time mortality fell sharply, and a sizeable proportion (about a third) of the world population still have not at the time of writing, and the distribution is very uneven, so that in some countries, a large majority have not. Yet even in these countries, Covid-19 has ceased to be a Pandemic disease.

This does not amount to an argument against the effectiveness of Covid-19 vaccination for reducing mortality. Vaccines significantly reduced the risk of death from infection with severe disease. But so did the emergence of Omicron late in 2021, a milder variant of Covid-19. The vaccines still saved many lives, especially from earlier variants, notably Delta. But this was only in places where vaccination was rolled out early enough to do so. Where vaccination programmes did not succeed before the Delta wave ran its course, the end of the Pandemic was not brought about by any human intervention at all.

The vaccines and the emergence of Omicron both contributed to significantly lowering the ultimate pandemic death toll, as both reduced infection fatality rates significantly. Ultimately, infection fatality rates would have fallen without either vaccines or Omicron, because infection with Covid-19 confers long-lasting protection against future severe disease. When everyone has been infected at least once and survived, the risk of dying during a subsequent infection is an order of magnitude lower compared to the risk of dying from the first infection (Medić et al. 2022). Infection fatality rates would thus have fallen eventually, as the proportion of survivors grew. However, a significant proportion of first infections would have been fatal, even with Omicron. The success of Covid-19 vaccines was to increase the number of survivors of the first infection, and by doing so, there is no reasonable doubt that they saved many millions of lives. Vaccines were generally rolled out by vulnerability group, and contributed importantly to reaching a situation where first infections were among people who either had some immunity or, if not, were less vulnerable (especially, were young) enough for risk to be greatly reduced with much reduced mortality. They continue to contribute by protecting people who would be more vulnerable to severe Covid-19: older people, immunocompromised people. They also help to reduce infections among people in close contact with vulnerable people, although the effect on transmission is much smaller, as discussed below. The effect wanes quickly (<3 months) and is significantly smaller with Omicron compared to prior variants, such that it cannot be relied on by anyone. This means that those interacting with vulnerable people need to take other precautions to protect vulnerable people.

7.3 The Covid-19 Vaccine Dogma

It may sound perverse to say there was some sort of dogma around Covid-19 vaccines, if they saved millions of lives. It may also appear dangerous, since doing so might reduce vaccine uptake among people who would benefit. Even if there was exaggeration or over-enthusiasm in some quarters, is it really helpful to point this out? Did the exaggeration do any harm?

We believe that it is indeed important to identify the Vaccine Dogma and refute it. First, it has done harm. Second, it continues to do harm. Third, the dynamic of which it is a part – the dynamic of a scientific orthodoxy – is harmful. Fourth, it is better – we believe – to strive for an intellectual culture in science, academia, and society at large in which falsehoods are not accepted for their pragmatic convenience, and complicated truths are stated rather than avoided. Humans are routinely distinguished from other animals by their intellectual capacities; to treat each other as humans, we must trust each other to be able to understand. In the remainder of the chapter, we will explain why we say there is a dogma and how we think it is harmful.

The Covid-19 Vaccine Dogma is as follows.

> **Covid-19 Vaccine Dogma**
>
> Covid-19 vaccines are safe, effective, and necessary.

In its simplest and strongest form, the Vaccine Dogma (we take it as read that this refers to Covid-19) identifies *safety* and *effectiveness* with standards that are, in turn, endorsed by scientific authorities, and *necessity* with protection of both the vaccine recipient and others. Of course, variations exist. In this section, we show that the relevant official scientific organ of the US Government endorsed the strongest form of the Vaccine Dogma. The US Centers for Disease Control (CDC) is the body in question. It is clearly an important scientific authority, possessing high degrees of political and scientific status. The CDC was an important player in the Pandemic Orthodoxy. We will also identify some other bodies that adhered to the Vaccine Dogma, and contrast these with some that did not.

Until around May 2025, the CDC web page on Covid-19 vaccination stated:

> [Title:] COVID-19 Vaccines Are Safe, Effective, and Free
>
> [Body:] Everyone 6 months and older should get an updated COVID-19 vaccine.
>
> *(CDC 2023)*

The title obviously expresses the first two elements of the Vaccine Dogma. There is no mention of standards of safety or effectiveness; these attributes are conveyed as absolute. Screens in airport immigration halls convey the same message, too.

The necessity element is present in the claim that everyone six months or older should get the vaccine. This is also an absolute claim: the reader is not advised to get the vaccine if certain ends are desired or for certain reasons. It has the nature of what Immanuel Kant called a categorical imperative. "Should" has the moral force of "must," but lacks the corresponding legal backing. It is also a universal claim: with the exception of young babies, everyone should be vaccinated. Vaccination is safe, effective, and everyone (except small babies) should do it: safe, effective, and necessary.

As updated in July 2023, the page contained no other text content, only a large number of links to other resources. In April 2024, the following text was added.

> People aged 65 years and older who received 1 dose of any updated COVID-19 vaccine (Moderna, Pfizer-BioNTech or Novavax) should receive 1 additional dose of an updated COVID-19 vaccine at least 4 months after the previous updated dose. For more Novavax information, click or tap here [a link is provided].
>
> *(CDC 2024a)*

This additional content adds a further requirement for a sub-group, but does not modify the safety, effectiveness, or necessity claims.

The link indicated in the text points to a page titled "Stay Up to Date with COVID-19 Vaccines" (CDC 2024b). That page again recommends vaccines for everyone older than 6 months: one dose for everyone from 5 years and up, "multiple doses" for babies and children between 6 months and 4 years, and various possible additional doses for older and other vulnerable people.

Other parts of the site reinforce the message. The page titled "Safety of COVID-19 Vaccines" (CDC 2020) leads with a section titled "What You Need to Know," and it includes all three items of the Vaccine Dogma, even though only one of them concerns safety. The first thing you need to know is as follows:

> COVID-19 vaccines are safe and effective.
>
> *(CDC 2020)*

As one might expect, on this page, safety is linked to standards of safety and effectiveness. Vaccines undergo a rigorous approval process. However, the statement that they are safe and effective is still not relativized: there is no implication that, if other safety standards were adopted, the vaccines would be unsafe relative to those standards, nor likewise for effectiveness. These properties are conveyed as absolute. The standards do not define the properties, but state that there are rigorous tests for them.[1]

Necessity appears further down the list:

> CDC recommends everyone ages 6 months and older get an updated COVID-19 vaccine to protect against serious illness.
>
> *(CDC 2020)*

This time, the "should" claim is connected to the end of protecting against serious illness. However, this cannot be taken as a serious statement of rationale. It is the statement of a shared assumption, something already implied in this context. It remains reasonable to see this as a statement of unqualified necessity.

The remainder of the site contains the same message. There is no doubt that the CDC, a scientific authority, asserts the content of the Vaccine Dogma: that Covid-19 vaccines are safe, effective, and necessary. But did the CDC lay safety, effectiveness, and necessity down?

The Vaccine Dogma was not held universally among scientists or public health authorities, nor was it the only or even primary driver behind vaccination campaigns. At certain times, it was extremely powerful, and although its power has waned considerably, it remains powerful at the time of writing. Yet there were scientific authorities that were more sensitive to evidence about safety, effectiveness, and necessity.

The UK government provides a good example of an authority *not* subscribing to the Vaccine Dogma. At the time of writing, the official government website states:

> If you're at increased risk of serious illness from Covid-19, getting a Covid-19 vaccine can:
>
> - help to reduce your risk of getting severe symptoms
> - help you to recover more quickly if you catch Covid-19
> - help to reduce your risk of having to go to hospital or dying from Covid-19
> - protect against different strains of Covid-19
>
> *(NHS 2023)*

From (northern) spring 2024, vaccinations were offered to people over 75 years of age, who have a "weakened immune system," or live in a care home for older adults. Vaccination is offered free through the National Health Service.

The European CDC and the World Health Organization were likewise considerably more moderate in their recommendations and statements about Covid-19 vaccines. The US was by 2024 an outlier. The difference is not simply a difference of scientific opinion or policy approach. There is a difference between a scientific dogma and a non-dogmatic representation of and response to the scientific evidence. The Vaccine Dogma was much more influential at some stages in the Pandemic, and we now turn to the task of illustrating what made it dogmatic: its inappropriate maintenance in the face of recalcitrant evidence.

7.4 The Healthy Vaccinee Bias in Scientific Publications

A scientific dogma resists recalcitrant evidence. We now turn to illustrate this resistance, starting with the *effectiveness* component of the dogma.

One of the tools used by the Vaccine Dogma was the *healthy vaccinee bias* (or *effect*). Covid-19 vaccination campaigns had major effects. Nonetheless, some assertions of effectiveness by some scientific authorities were dogmatic. Where the healthy vaccinee bias operates, even an ineffective vaccine – a placebo – may appear effective, as explained below. The healthy vaccinee bias thus effectively guaranteed the truth of the effectiveness claim (on the assumption that the vaccine was not actually *increasing* mortality among Covid-19 sufferers). And, where there was a real effect (as there commonly was), it made the effect large enough to put it beyond dispute. Overstatement and the bias that supported it were thus an important part of the process by which the Vaccine Dogma was maintained inappropriately in the face of recalcitrant evidence.

The healthy vaccinee bias (or effect, interchangeably) is a well-known problem for observational studies evaluating the effectiveness of any vaccination programme. The problem is that healthy people are more likely to be vaccinated, while people with existing health problems are less likely to be vaccinated. There are many reasons for this, including that the immune system is too fragile to risk it, that normal side effects such as a mild fever will prove dangerous in these cases, and that the person has the disease already. Mortality is therefore likely to differ between the two groups even if no vaccine or a placebo is administered, simply because one consists of people with serious health problems. It is therefore difficult to distinguish the benefit of the vaccine from the effect of separating out the unhealthy people into a separate group.

If the healthy vaccinee bias is not acknowledged and dealt with, then the effect of a vaccine will be significantly overestimated by an observational study. This can be illustrated with an imaginary, ineffective vaccine – a placebo. Imagine that 1,000 people are offered the "vaccine" (placebo). Imagine that 10 of these people are seriously ill already (from something, it does not matter what), so do not receive the vaccine: call this the unhealthy unvaccinated group (UUV). Imagine, further, that 90 healthy people refuse the vaccine: call this the healthy unvaccinated group (HUV). Now suppose that the following (unrealistic) mortalities occur in the follow-up period. Ten per cent of the entire healthy population dies, distributed regardless of vaccine status (since after all, it is an inert placebo), and all ten who are already seriously ill (UUV) die.

$$\frac{(10\% \times HUV) + UUV}{HUV + UUV} = \frac{(0.1 \times 90) + 10}{90 + 10} = \frac{19}{100} = 19\%$$

Thus, the mortality in the "unvaccinated" group is 19%. Meanwhile, the mortality in the vaccinated group is 10%: this is just the mortality among all healthy people in the population, regardless of vaccine status. Yet if we ignore prior health status and take the unvaccinated group as a whole, its mortality is nearly twice that of the "vaccinated" group – that is, the group that received the placebo. Who knew that sugar pills could be so effective?

Where a real vaccine is administered in place of a placebo, and even where it is an effective vaccine, the bias still operates for the same reasons. This does not negate the effectiveness of the vaccine, but it makes it difficult to tell just how effective a vaccination programme is, through observational studies.

The healthy vaccinee bias does mean that it is extremely difficult to assess vaccine effectiveness. It does not afflict randomized controlled trials, where assignment to treatment and control groups is random and not associated with health status. However, by their nature, trials cannot assess the effectiveness of specific vaccination programmes (unless, of course, those are part of the trial). To find out what effect an actual vaccination programme was, we need to try to work out what would have happened if that very programme had not been implemented. To do this, we need to take a range of information into account. One kind of information is that provided by observational studies comparing the incidence of outcomes of interest among vaccinated people and unvaccinated people. But there are others, notably the effect size observed in trials of the vaccines. If these two numbers are very different, then we need to think about why, and we cannot simply assume that the observational study provides an accurate estimate of the effect of the programme. We must also include methodological considerations such as the healthy vaccinee bias. In short, we must apply our minds.

The healthy vaccinee bias was, unfortunately, common even in prestigious journals. One study in the *New England Journal of Medicine* showed a 90% reduction in Covid-19 mortality for the boosted population (Arbel Ronen et al. 2021). However, a responding letter further showed that there was an even larger reduction (94.6%) in non-Covid-19 mortality in the boosted population (Høeg, Duriseti, and Prasad 2023). Remarkably, this was not disclosed at the time of publication, since all-cause mortality information was not included (a fact also noticed by another letter writer (Rohban 2022)). Such a difference shows that some factor besides the vaccine is operating to give rise to differences between the boosted and unboosted populations. This does not exclude an effect of vaccination, but it completely punctures the inference drawn in the published study, which implicitly depends on vaccination being the only reasonable explanation of the difference.

This is one example of many papers that almost certainly suffer from a healthy vaccinee bias. Any paper showing a high effectiveness against mortality (e.g. 90%) of a subsequent dose compared to a preceding dose, most likely suffers from a healthy vaccinee bias for three reasons. First, the nature of Covid-19 immunity from any source is that protection against serious disease is long-lasting, meaning that the scope for boosters to achieve major reductions in mortality is small. Second, the efficacy of boosters in trials (which are less likely to be subject to healthy vaccinee bias) is low, presumably for the same reason (Thomas et al. 2021). Third, from a theoretical perspective, the healthy vaccinee bias is likely to compound with each booster. With each subsequent dose, any individuals who have developed further health issues will be removed from the group receiving the next dose. The group receiving a third or fourth dose will thus have been whittled down to a core that is likely to be considerably healthier than people not receiving the booster. So we have three considerations that triangulate to make a 90% effect of boosters on mortality implausible: the evidence from trials, the intrinsic nature of Covid-19 immunity such that protection against mortality wanes slowly, and the theoretical argument for the compounding of the healthy vaccinee bias with each boost. This kind of triangulation is at the heart of good epidemiological reasoning, and no amount of statistical analysis can replace it (Vandenbroucke, Broadbent, and Pearce 2016; Broadbent, Vandenbroucke, and Pearce 2016).

7.5 Healthy Vaccinee Bias in Public Communications

In public communications, the healthy vaccinee bias gave rise to polemicism and visual misrepresentation. Several graphics (perhaps aspiring to be memes) circulated concerning vaccination status among hospital admissions. Obviously, admission to the hospital after receiving a vaccine is strongly influenced by your health state prior to being offered a vaccine. This remains true even if you are admitted with Covid-19: your prior health state may have compounded the threat infection poses to you, or may not even be the reason for your admission at all. Graphics such as the examples in Figure 7.5 and Figure 7.6 do not discriminate between people who were not vaccinated because they were not healthy enough

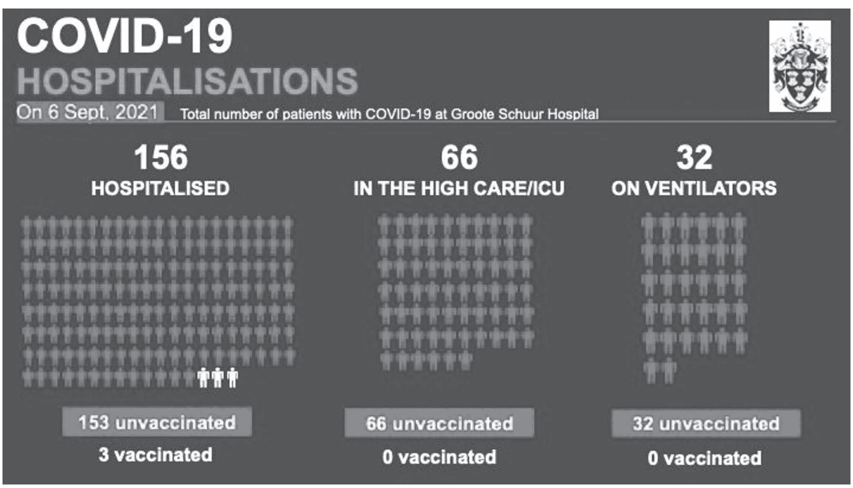


FIGURE 7.5 A graphic afflicted by the healthy vaccinee bias

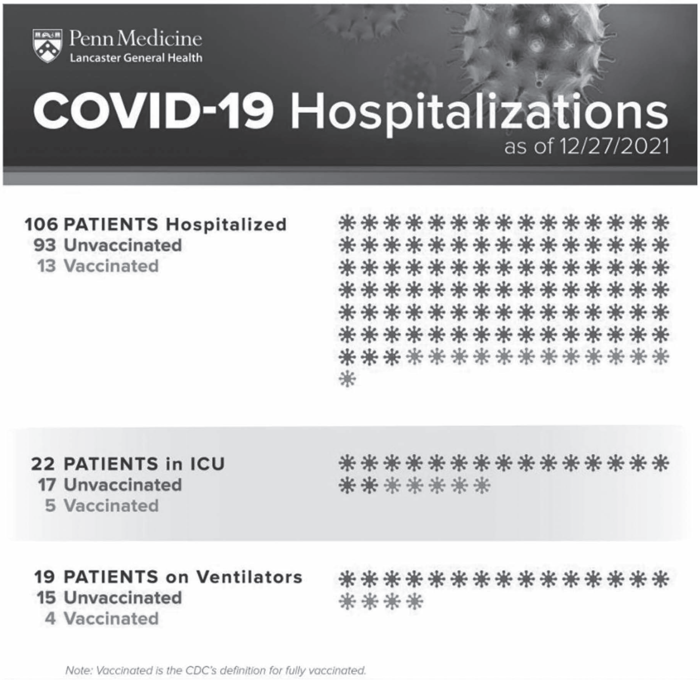


FIGURE 7.6 Another casualty of the healthy vaccinee bias

to be vaccinated and healthy people who were not vaccinated. It is therefore impossible to tell from these figures what effect Covid-19 vaccines are having.

Even if it is reasonable to simplify for purposes of public communication, there remains a difference between simplification and misrepresentation. These graphics do not convey a simplified version of the truth. They convey a falsehood.

It may be replied that the graphics do not explicitly *assert* that the difference between the groups is entirely due to vaccination status. However, the implication is so strong that one cannot regard them as anything other than misrepresentation. There can be no other reason to present graphics such as these than that the viewer draws the inference that vaccination is responsible for the depicted differences in mortality. That is because the graphics are meaningless without information about underlying health conditions. Probably, an effect would still be visible with the correction, albeit smaller, but the graphics do nothing to support this claim, let alone quantify the effect. Indeed, they are compatible with a negative effect of vaccination, obscured by a larger healthy vaccinee bias. Once the healthy vaccinee bias is considered, one sees that the graphics convey no definite information at all about the effect of the vaccine, and thus, there is no point in creating a graphic of this kind without considering the effect of this bias.

Other biases operate too in these simple graphics. In particular, there is a strong association between socio-economic status (including factors such as income and education level) and vaccination. Poorer people are less likely to be vaccinated. There is also a strong association between socio-economic status and health outcomes in general, as well as Covid-19 mortality in particular. This makes it more likely that an unvaccinated person is less healthy than a vaccinated person. This means that the healthy vaccinee bias is more than just a theoretical possibility; it is probable that it operates in circumstances such as those shown in Figure 7.5 and Figure 7.6. The people indicated by red symbols on both graphics are almost certainly much less healthy, as a group, than the general population.

These biases show that the effectiveness component of the Vaccine Dogma was maintained in the face of recalcitrant evidence, in a broad sense of that phrase. The evidence – the actual scientific evidence – was not recalcitrant in the sense of being *contrary* to the assertion of effectiveness. But it supported a *complex picture of effectiveness*. In this sense, it was recalcitrant. The Vaccine Dogma makes a simple and categorical assertion. The evidence does not support that simple, categorical assertion, as regards effectiveness, and the evidence presented in its support was sometimes subject to biases that would have made a placebo look like a wonder drug. Vaccines were indeed effective, but the effect was overstated and insulated from evidence, instead of being accurately presented on the basis of the most reasonable interpretation of available evidence.

7.6 Absolute Statements of Safety

We turn now to the other aspect of the Vaccine Dogma: safety.

Vaccines are administered to healthy people. Public tolerance of adverse reactions is very low. Yet unless a vaccine is completely safe, any large campaign will kill some people who would not otherwise have died, for the sake of saving the lives of a much larger number of people. This is accepted in certain situations, like war. Nobody tries to tell anybody that deployment on the front line of an armed conflict is safe. Yet this language is sometimes used for vaccination. Risks may be very low, but the reality is that any vaccination programme will cause some deaths even if it saves many lives. Moreover, if there are multiple vaccines, and one has higher serious adverse events, then these events are avoidable. It is therefore critical that serious adverse events are closely and accurately monitored.

Some people are inclined to see "safe" as a misleading or even deceptive term for this reason. They may oppose all vaccinations because they regard the harms as unacceptable. This position is often cast as irrational, but in fact, sacrifice for the greater good is a topic on which opinions may reasonably differ. Whatever their position in theory, in practice, not even a utilitarian would accept a vaccination that killed just one fewer person than it saved.[2] On the other hand, most people have a strong intuition that to kill someone innocent is always wrong, no matter how many people are saved. For some people, this is much more vivid in the case of vaccination than for others. Political debate about vaccination has many aspects and is often unpleasant. But at least as regards safety, it is important to recognize that even one death from a vaccination campaign that would otherwise have been avoided provides legitimate grounds for reasonable disagreement.

However, in the context of public health and public policy more generally, utilitarian thinking tends to prevail. Trade-offs are seen as an inevitable part of nearly any policy decision. Permitting motor vehicles to travel at high speeds implies killing some people so other people can get around quicker. From this perspective, vaccination presents no special challenges. People with this mindset are inclined to understand the absolute word "safe" as a reasonable way of indicating that the risks, while present, are low enough for a reasonably prudent person to accept. This is the sense in which we say air travel is safe even though aeroplanes crash.

Our argument is that even if one accepts this perspective, there was dogmatism around the safety of Covid-19 vaccination: safety was inappropriately maintained in the face of recalcitrant evidence. We do not say that the vaccines are *not* safe. Rather, we argue that safety considerations were misrepresented and that dangers were downplayed. This was part of a wider lack of nuance, resulting in – or represented by – crude implementation policies. The problem is not the vaccines themselves, but the way they were promoted and implemented.

One way in which dogmatism arises is in statements that are absolute or lack appropriate qualification. Statements that the Covid-19 vaccines are *absolutely* safe, we would regard as dogmatic, since they stifle discussion of *how* safe a vaccine is, or whether a vaccine is safe *enough*. No vaccine, indeed no medicine, is absolutely safe. Haste and the difficulty of public communication may provide mitigating circumstances in many contexts, but certainly not in all. The CDC again provides a good example, repeatedly describing Covid-19 vaccines as "safe" with no qualification. There has been ample opportunity to nuance and correct any hasty statements, because more safety information is now available, and because the urgency is past, making the CDC's persistence dogmatic.

The dogmatism is also damaging, not to mention embarrassing, when further research makes it impossible to ignore recalcitrant evidence. This happened in relation to early statements that masks were not effective against the spread of Covid-19 (ironically), and it happened again in relation to vaccines. Safety concerns emerged soon after Covid-19 vaccination programmes began, and have continued to emerge since.

In early 2021, evidence emerged linking the AstraZeneca vaccine to extremely rare but serious (including fatal) blood-clotting events (European Medicines Agency 2021). The AstraZeneca vaccine was the first vaccine to be discontinued by several countries (e.g. Australia, South Africa) that had based their strategy on this vaccine, resulting in significant delays in vaccine rollouts as new vaccines were sourced.[3]

Shortly afterwards, in April 2021, Reuters reported cases of myocarditis in young men in Israel after receiving the Pfizer-BioNTech Covid-19 vaccine (Reuters 2021). By 2022, there was little doubt about a causal link, with a myocarditis hospitalization risk for adolescent boys of approximately 106 per million after the second dose of the Pfizer vaccine (Krug, Stevenson, and Høeg 2022). Roughly this level of risk has previously provoked the withdrawal of a vaccine for rotavirus.[4] These concerns gave rise to the concerns already discussed about cost–benefit ratio of a second dose of the Pfizer vaccine for healthy adolescent boys, and ultimately to policy changes in many places.

Levels of myocarditis were even higher after the Moderna vaccine (Patone et al. 2022), and in October 2021, some Nordic countries restricted the use of the Moderna vaccine for those under 30 years of age, recommending Pfizer instead (Taylor 2021).

The Johnson and Johnson vaccine was also later discontinued in some countries due to similar extremely rare but serious blood-clotting disorders (ten in a million in women aged between 30 and 49), with nine fatalities in 18 million people. These, too, were raised in 2021. However, some countries continued with the vaccine, but discontinued it when further concerns emerged – notably the CDC. This decision drew criticism even in 2021 (Pegden 2021), and the change of position in 2023 raises the question of whether those authorities that persisted despite the safety concerns should have responded differently to the earlier advice, and whether they should have made the safety information more explicitly available to vaccine recipients. While these levels of adverse event fatalities might have been acceptable if these were the only vaccines available, the availability of other safer alternatives made the continued use of these vaccines much more questionable.

To be clear, this summary does not show that Covid-19 vaccines are unsafe, nor that they ought not to have been used. To reiterate, we believe that Covid-19 vaccines saved millions of lives and that in certain situations they had their place: that the risks and the harms were worthwhile for the benefits. What this summary shows is that absolute statements of safety are incorrect. On the contrary, since very early on, there has been a considerable quantity of safety information available. Some authorities, commentators, and opinionated individuals adjusted their positions accordingly as this information emerged. Others, however, did not.

The Vaccine Dogma is particularly evident when absolute assertions of safety are not accompanied by adequate indications of this information. Again, the US CDC is the paradigm example, because it persisted so long in a stance that was abandoned much sooner in most other places. The CDC's page on the safety of Covid-19 vaccines said nothing at all about risks or adverse events in its "What

You Need to Know" section, implying that you do not need to know about them. Further down, there was a section headed "Adverse Events (Health Problems) Are Rare," but this section clearly minimized the risks. It pointed out that an adverse event "can be caused by the vaccine or can be caused by a coincidental event." It provided a hyperlink on the words "adverse events" that pointed to a generic page about adverse events, not to information about adverse events after Covid-19 vaccination. In fact, there was no information at all on the page about what adverse events might occur, or about where to find reliable information about adverse events. A hyperlink was provided on the words "benefits of Covid-19 vaccination" but not on "outweigh the known risks." If somebody really wants information about the safety of Covid-19 vaccination, they will not find this page helpful.

7.7 Individual and Group Variation

Absolutism about vaccine safety is particularly problematic when there is variation in safety between groups, coupled with the highly differing benefits offered to those groups. Safety information about Covid-19 vaccines was distorted by presenting the overall benefit against the overall risk, without appropriately stratifying different groups. When everyone is lumped together, the dangers of vaccination pale in comparison to the overall benefit. However, the risk of Covid-19 differed by orders of magnitude between the oldest and the youngest, and also significantly between those with co-morbidities and the healthy. The benefit of a first dose and subsequent doses may also be expected to differ, given that immunity against severe disease wanes slowly, while the risks may – or may not – be greater.

Some authorities, including the CDC, continue with over-general assertions that the benefits of Covid-19 vaccination outweigh risks, without differentiating between groups and doses. When the CDC met to consider the risks and benefits of vaccination in June 2021, it lumped first and second doses together (Pegden 2021). But this is too simple. Both sides of the risk–benefit ratio vary hugely between individuals. In particular, the risk of cardiac complications is particularly high in boys. The CDC reported 63 per million in 12–17 year olds and 50 per million in 18–24 year olds, both above public tolerance for adverse events in previous episodes that had led to the withdrawal of vaccines. This risk is much lower in older age groups (and females), while the benefit to older groups is high. Yet the CDC recommended, not only first-time vaccination for 12–24 year old males, but multiple doses. Based on the information available (at the time) about persistence of immunity against serious disease, this recommendation is seriously flawed.

A stratified risk–benefit analysis that also separated first and second doses was not done until 2022, and then not by the CDC (Krug, Stevenson, and Høeg 2022). It found there was still a benefit to vaccinating children in relation to Delta, but that the benefit from boosters was much less clear, and that there was no apparent benefit in relation to Omicron.

Wider costs were not considered, and it has been argued that when they are, vaccinating children does not tip the balance:

> Implementation of a large-scale immunisation programme, however, comes with both financial and opportunity costs – for example, diversion of health-care staff and resources could potentially affect the provision of other crucial health-care services, such as routine childhood immunisation programmes. Clinicians and parents must balance the relatively small risks of severe disease outcomes with the relatively small risks that accompany vaccination in children aged 5–11 years.
>
> *(Ladhani 2022, 75)*

These authors conclude much as we do, that simplified messaging around Covid-19 vaccination is unhelpful:

> Above all, public messaging of the risks and benefits of vaccinating children against COVID-19 needs to be clear to encourage public confidence in vaccines and trust in those advocating for vaccination to prevent other, more serious diseases
>
> *(Ladhani 2022, 75)*

From 2022 onwards, almost all countries recommended the vaccine only for high-risk groups. But there was a period of time when the current was flowing the other way, and there remain influential authorities with a dogmatic stance on Covid-19 vaccines, including their safety.

As usual, we must emphasize that none of this means that Covid-19 vaccines are unsafe. To say that would be to defeat our main point: that there is a need for nuance and critical thought. In particular, it is far from clear that Covid-19 vaccines are safe *enough* for young people, relative to the benefits and the other costs of vaccinating. We suspect not, but you might reasonably disagree. Our point is that it is a debate that may reasonably be had. What we oppose is maintaining that Covid-19 vaccines are categorically safe. Given the evidence, it is inappropriate to maintain this. But the opponent is rhetorically pushed towards saying they are *not* safe, an even less reasonable position. The Dogma establishes a false dichotomy and shuts down discussion of real issues: notably, risk–benefit for different groups, and acceptable levels of risk.

7.8 Lack of Acknowledgement of Missing Data

We would also include as dogmatic misrepresentations of the extent of safety data available. To present a vaccine as safe before enough people have taken it for adverse events to be detectable is obviously a departure from evidence.

Large studies are required to detect adverse events from vaccines because public tolerance is low, meaning that even a rare adverse event will often result in public outrage. Rotashield, developed against the rotavirus, had an estimated risk of causing intussusception of 100 in a million and was withdrawn within a year after its introduction. Pandemrix, developed for the H1N1 pandemic, was estimated to increase the incidence of narcolepsy by 40–90 per million, and its use was widely restricted. Denvaxia, developed against dengue fever, exacerbated dengue fever infections in some individuals who had never been infected by dengue fever before and was withdrawn after 2 years in 2017.

Covid-19 vaccines were developed and received emergency use authorization in a record time of less than 12 months during the pandemic. Under normal circumstances, the entire process from preclinical development to regulatory approval often takes between 10 and 15 years. This was widely celebrated as a great achievement. At the same time, there is obviously no way that a trial so short can detect longer-term negative effects. There is also the possibility of adverse events too rare to have been detected in those studies. There is a reason the usual process takes so long, beyond lack of urgency. Emergency approval is a perfectly reasonable process, of course. However, it is not as reasonable to say, in 2020, that safety has been established, either in absolute terms or in relation to the safety standards applied to vaccines generally. There was simply no way that these safety standards could have been satisfied so soon. Rare adverse events or those emerging after a longer period of time cannot be discovered in studies that are not large enough or long-lasting enough. Again, to present something as absolutely safe in this context is dogmatic.

The safety of Covid-19 vaccines was maintained by some scientific authorities in the face of recalcitrant evidence. But it is not accurate to say that Covid-19 vaccines are safe, categorically, without some reference to the complicated evidence on the matter. The unqualified assertion that Covid-19 vaccines are safe creates a problem for someone who wants to disagree, yet does not want to make an equally blunt and even less desirable statement that the vaccines are not safe. Debate in the middle ground is shut down by dogmatic assertions of safety. Yet it is the debates in the middle ground that lead to sensible policy decisions and that treat citizens with appropriate respect, the kind that engenders trust. On vaccine safety, as on other matters, scientific dogma does not facilitate good vaccine policy.

7.9 Who Should Be Vaccinated?

We turn now to consider the necessity of vaccination against Covid-19. Just because a pharmaceutical intervention is effective at achieving some desirable end does not mean it is necessary for anyone, much less for everyone. Whether it is necessary depends on whether the end itself can be considered so important as to be called necessary, and this, in turn, depends on the type and scale of the benefit offered. In this section, we distinguish the different kinds of benefits offered by Covid-19 vaccination, and in the remainder of the chapter, we consider to whom these benefits accrued and what the implications are for the "necessity" of vaccination, especially for those to whom vaccines offer fewer or smaller benefits.

Getting vaccinated against any disease offers an individual two potential benefits: one to themselves, and one to other people. Each kind of benefit subdivides further. For the individual, the two potential benefits are against infection and against severe disease or death. For other people, the two benefits are the reduction in the chance of infection proportional to how many other people have been vaccinated and the effect of crossing a herd immunity threshold. To say that a vaccine is necessary must be to say that it is necessary to achieve – or at least raise the probability of – one or more of these potential benefits. In the case of Covid-19 vaccines, the distinction between these benefits was not made. Yet Covid-19 does not offer all of these benefits. Therefore, the necessity was commonly misstated.

The fundamental principle of vaccination is to teach the immune system how to respond to an infectious disease without exposing it to a serious threat. The immune system responds to viruses (including Covid-19) by creating antibodies and activating T cells, both of which are critical for neutralizing or destroying the virus. When encountering the virus for the first time, the immune system must figure out how to generate these responses. However, if it has encountered the virus before, it "remembers" through memory B cells and T cells, allowing for a faster and more efficient response.

Vaccines introduce antigens to the body that mimic the pathogen, which could be inactivated viruses, live attenuated viruses, or specific proteins characteristic of the virus. If a vaccine works as it should, it will provoke the immune system to manufacture antibodies and activate T cells that would target the virus. The vaccine is not dangerous, so the immune system has time to create these defences without the risk of severe disease. When the real virus arrives, the body can respond quickly enough to prevent severe illness. Modern vaccines, like mRNA vaccines, provide genetic instructions (mRNA) for cells to produce viral proteins, which then elicit an immune response, creating immune memory without significant risk.

Vaccinations for different diseases have different effects for various reasons. One concerns incubation. Many viruses replicate in specific locations in the body before spreading. For example, measles incubates for 11–12 days, often in the lymph nodes. If the immune system clears the virus during this period, serious symptoms do not appear. Although transmission of measles can occur during the late incubation period, even before symptoms like the rash emerge, the longer incubation period gives the immune system ample time to mount a response, even in situations where only a small amount of immune memory remains, significantly reducing transmission risk.

Covid-19, like many other respiratory viruses, replicates initially in the mucosal surfaces of the nasal passages. Its incubation period is approximately six days, but it can vary. During this period, the host may be infectious, as the virus can spread from the mucosa to others even before symptoms appear. The immune response, including antibody production, can begin at these mucosal sites. If there are already antibodies and memory T cells present from a recent infection or vaccination, the body can often mount a sufficient response to reduce transmission risk; however, as antibody levels wane, transmission risk increases.

This explains why vaccination against some diseases, such as measles, can offer long-lasting protection, while vaccination against others, like Covid-19, offers only moderate and short-term protection against infection, but still provides good long-term protection against severe disease.

These characteristics also explain why lasting herd immunity against Covid-19 cannot be achieved in practice. Different diseases have different herd immunity thresholds; this is the proportion of the population that needs to be non-susceptible to infection in order for the reproduction number to fall below 1.0, that is, for incidence to start reducing. For Omicron, using the standard SIR formula and values of R in the range 5–9.5 (Liu and Rocklöv 2022), this threshold is estimated to be in the range of 80–90%. The implication is that if a vaccine were 100% effective, then only 80–90% of the population would need to be vaccinated. However, the effectiveness of available Covid-19 vaccines against Omicron is well below 80%, and by the time the last people receive their vaccination, effectiveness among the first-vaccinated would have dropped to around 20%. Clearly, achieving herd immunity via vaccination is completely impossible for Omicron (and was also impossible for Delta).

Herd immunity via infection is possible, but it is transient. This effect largely explains why individual waves came to an end during the pandemic. Over time, immunity against infection wanes, resulting in subsequent waves (even in the absence of new variants or seasonal changes). While one population is recovering from infection, the virus is circulating elsewhere. Unless the immune population is truly isolated – like a permanent population in a sealed space station – there will be renewed contact with the virus, and it will once again be able to infect people. During Omicron, waves were as little as three months apart in some countries due to the rapid emergence of new variants (Chatterjee et al. 2023). Thus, lasting herd immunity is not a practical possibility, whether by vaccination or infection.

To summarize, Covid-19 vaccines, at the time of writing, offer the following benefits.

1 Long-term, significant individual benefit against serious illness and death.
2 Short-term, moderate individual benefit against infection and mild illness.

3 Short-term, moderate population benefit to the wider population against infection, proportional to the number of people recently vaccinated or boosted.
4 No population benefit for reaching a herd immunity threshold, which at the time of writing cannot be reached for Covid-19 by either vaccination or infection.

7.10 Overreach, Mandates, and Kale Salad

Vaccination campaigns and policies commonly ignored the limitations of Covid-19 vaccination. Specifically, they often misrepresented the benefits.

Individual protection against severe disease was a benefit of vaccination against Covid-19, but this benefit is obviously proportionate to the risk of severe disease. Promoting vaccination among low-risk groups on this basis commonly involves overstating the benefit. Even if safety is not a concern, the benefit of being vaccinated may be very low. Urging a parent to vaccinate their child against Covid-19 for the child's safety may imply a much higher level of risk from non-vaccination than is real.

Similarly, for many adults in younger age groups, the threat posed by Covid-19 was small. Urging people to receive a vaccination may have implied a much higher level of risk and thus benefit than was real, especially by implying that the risk was larger than other risks on which much less public health energy and money is spent.

However, more unfortunate overreach occurred in relation to protection against infection, and the corresponding benefits at the population level of transmission prevention. These were often significantly misrepresented, resulting in policies that were ill-advised and unjust. For example, in December 2020, in an interview with CNBC, Dr Anthony Fauci stated that if 70%–80% of the population were vaccinated, the virus would have nowhere to go. It is difficult to see why he would have said this, because he also admitted he was uncertain how long immunity would last. A well-prepared interviewer from *The New York Times* put pressure on this point later, asking if it was reasonable ever to have hoped that a vaccine against Covid-19 could confer herd immunity. Fauci said that in 2020, "it was not completely out of line to think that you could get protection against infection." But the more important question is whether it was out of line to assert that herd immunity was achievable.

It was certainly uncertain, and probably improbable, that Covid-19 vaccines would offer herd immunity, given what was known about respiratory viruses (including other coronaviruses and influenza). But it was also politically and socially irresponsible to assert that they would.

The prospect of herd immunity led many authorities to exert pressure on individuals to get vaccinated. In some places, mandates were enforced. For example, in Austria, a law was passed in January 2022, requiring all residents aged

18 and older to be vaccinated against Covid-19. It included a timeline where non-compliance would initially result in warnings, followed by fines. The mandate excited considerable controversy and was lifted on 23 June 2022. Many other countries applied very significant pressure short of mandates, for example, permitting certain activities or access to certain spaces only on certification of vaccination status.

In the private sphere, some employers insisted on vaccination of employees, also supported by governments issuing official vaccination status certificates. In many places, educational institutions insisted that their students be vaccinated. Herd immunity provides a powerful reason for states to pressure their citizens, as well as employers to pressure their employees and individuals to pressure each other. Liberty, jobs, and friendships are lost to this kind of pressure. Vaccination is often portrayed as "doing the right thing," because vaccination against diseases like measles offers benefits to one's contacts and, through herd immunity, to the population as a whole.

But what is the rationale for widespread vaccination? If it is for widespread individual protection to the recipients of the vaccine, then this sort of pressure is a huge overreach. As Stephen John points out, mandating a Covid-19 vaccine is on a par with mandating kale salad: beneficial to the individual, but unjust nonetheless (John 2022). The underlying assumption, shared across vaccine mandates in general, is that vaccination offers benefits extending beyond the individual recipient. It was this wider good that justified authorities – in their eyes, who required unwilling individuals to be vaccinated against Covid-19. But the immunological characteristics of Covid-19 itself, and the properties of current vaccines, render this argument completely wrong.

There is not even a strong altruistic argument for vaccination as a protection strategy for people at high risk. Imagine a university, where most students are at low risk of serious disease, but members of staff are at higher risk – some a lot higher. If herd immunity is impossible, as is the case with Covid-19, and those at higher risk are vaccinated themselves, there is very little additional benefit from increasing vaccination rates amongst students. The virus will continue to spread, perhaps a bit slower in a recently vaccinated population, but ultimately, everyone will be infected by Covid-19 on a regular basis. It would then not take long before the most relevant immunity will be from the last infection, and not the original vaccination. This, of course, is exactly what happened during the Pandemic. The real benefit of the vaccine is not to the wider population, but to the person taking it, since it offers effective protection against severe disease, but is not very effective at preventing transmission.

Mandates thus were not justified by any principle of democratic society, whether implemented by states or within private companies, and the level of pressure that was applied in states where mandates were not implemented was also unwarranted. The clamour of many commentators in the media and opinionated citizens at large was also unwarranted. There was neither a scientific nor a constitutional basis for such a strong vaccination drive.

A particularly striking illustration of the operation of a dogma comes from reactions to increased knowledge about waning immunity. As the vaccines were rolled out in 2021, it soon became clear that there were a significant number of breakthrough infections and that vaccinated people could spread the virus. We have already seen how waning immunity makes herd immunity unachievable for Covid-19. Yet this was seen by some as strengthening the rationale for mandates. A higher threshold simply meant that efforts to reach it should be intensified. Yet we have already seen that the threshold was unattainable. On 18 March 2021, a news article in *Nature* had already pointed out that herd immunity was impossible for Covid-19 (Aschwanden 2021).

Similar remarks apply to the vaccination of children, who, with the exception of a vulnerable few, very rarely suffer serious or fatal Covid-19. If people catch Covid-19 as children for the first time, and continue to be exposed to it on a seasonal basis, then they will probably retain immunity against serious disease for the rest of their lives, at no financial cost. As well as confusing "the public," Covid-19 vaccination also caught the theoretical literature on vaccination off guard. The ethics and politics of vaccination have received extensive attention in philosophical, historical, bioethical, and other academic literature (e.g. Holmberg, Blume, and Greenough 2017; Goldenberg 2021). Much of this work focuses on reasons for hesitancy, public trust, and the social contract surrounding immunization. An underlying assumption is that vaccination offers benefits to others, as well as the person being vaccinated (Flanigan 2014). There is an acknowledged puzzle around cases where individuals do not themselves benefit (Rose 1992), and in extreme cases cannot, as when vaccinating both girls and boys against HPV to prevent cervical cancer. However, the Covid-19 case presented a novel configuration: vaccines that offered meaningful individual protection against severe disease but limited and short-lived effects on transmission. This distinction undercuts the conventional ethical rationale for coercive vaccination policies, which typically rely on robust indirect protection (herd immunity) to justify mandates. The legal enforcement of vaccine mandates relies on the idea that "Vaccine refusal harms and risks harming innocent bystanders" (Flanigan 2014, 5). But in the case of Pandemic-era Covid-19 vaccination, this is not true. Arguments for coercion based on third-party protection simply do not apply.

7.11 Financial Bias, Boosters, and Children

It is uncontentious that studies conducted by people with a financial interest in a positive outcome are more likely to have a positive outcome than "independent" studies, that is, ones conducted by people without a financial interest in any outcome (Lexchin, Bero, and Djulbegovic 2003). There are many possible mechanisms, ranging from outright fraud at one extreme to unconscious over-enthusiasm at the other. For present purposes, it does not matter. The point is that financial bias is a possibility whenever there is a financial interest in a particular outcome.

The financial incentives relating to Covid-19 vaccination are huge, and so, therefore, is the potential for financial bias. Specifically, the elements of the Dogma are all subject to bias. There are financial interests in the safety, effectiveness, and necessity of Covid-19 vaccines. This means that research relating to each of these is subject to potential bias.

The potential for bias is not the same as its reality. It is impossible to entirely separate financial interests from biomedical research because the sums of money involved are only available through investment, which means that a return is required. For this reason, mechanisms exist within biomedical research that are intended to mitigate the effects of financial bias. Many of the mechanisms of evidence-based medicine are designed with this in mind. One of the simplest is that financial interests are required to be declared. This declaration is made when results are published, so that anybody seeking to rely on the results will know that there is a potential for financial bias and will be able to take this into account when using the findings.

However, financial interests are not declared when medications become the subject of a public health campaign. This is true of vaccination programmes in general. The fact that pharmaceutical companies make very large amounts of money from routine vaccination is not conveyed to parents making a decision as to whether to vaccinate their children. Indeed, the process is not generally conveyed as a "decision" at all, but as a routine part of responsible child-raising. The question of financial bias is left to people with very strong views opposing vaccination, and to raise the question of whether vaccination programmes are driven in part by profit is often construed as expressing an unreasoning scepticism: in short, as a sign of being an "anti-vaxxer."

Questions about public health messaging, and of paternalism in public health, are very complex. There are a number of situations where conveying full information might be considered (by some, at any rate) unnecessary or wrong, from a public health perspective. It may be that vaccination is one of them, and that potential financial bias is one of the things that ought not to be conveyed. One argument, for instance, might be that by merely conveying the *potential* for financial bias, there is an implication of the *reality* of financial bias. Assuming that financial bias has been controlled and ruled out, this would be an incorrect implication. Another argument is that the people encouraging uptake do not themselves profit, rendering financial bias less likely; indeed, wide uptake of a vaccine actually costs the public health authority money (assuming the vaccine is publicly provided). A third argument is that reasons can be disconnected from benefits: a gym urges you to join because it wants your membership fee, but your health still benefits if you go to the gym. So even if vaccination campaigns are motivated in part by profit, this does not matter, since their outcomes are still beneficial. In each case, raising the possibility of financial bias might be seen as unnecessary or wrong, due to the risk of reducing uptake for a bad reason.

These arguments are all highly contestable from an ethical point of view. But even if one accepts them for routine vaccinations that have been in use for a long time, there is still a good argument that the financial interests in relation to Covid-19 vaccination were undercommunicated. There are many different *contexts* for communication. At one end are the formal conflicts of interest statements on scientific publications. No reasonable person will deny that these are necessary, even if strongly inclined to favour paternalism in public health. The foregoing arguments against communicating a possibility of financial risk are not engaged in that context. At the other end of the scale, one has the medical professional holding the needle. No reasonable person will insist that this professional raise the possibility of financial bias before inserting the needle. The final checks should concern risk factors for serious reactions – allergies, complications, etc. – but not awareness of the risk that safety, effectiveness, and necessity may have been misstated due to the fact that the vaccine manufacturer profits from the upcoming injection. The foregoing arguments against communicating a possibility of financial risk would surely be engaged in that context. But there is a large space in between these two contexts: the space in which public debate occurs. It is in this space that we argue that the financial bias in relation to Covid-19 vaccines was underplayed.

For example, on 28 July 2021, CNN reported:

> A third dose of the Pfizer/BioNTech Covid-19 vaccine can "strongly" boost protection against the Delta variant – beyond the protection afforded by the standard two doses, new data released by Pfizer on Wednesday suggests.
>
> *(Howard 2021)*

The phrase "new data released by Pfizer" was a hyperlink pointing to Pfizer's financial report to investors for the second quarter. A financial report to investors is obviously subject to the potential for financial bias. Moreover, there is no statement of conflict of interest, as there would be in a scientific publication. Indeed, there *is* no conflict, because the interests of the management and the investors are obvious, and are aligned: they are the financial performance of Pfizer. The new data is indicated as "pending publication," which means that there has not yet been a process that might have removed this bias. It is true, of course, that the report is linked as a source. But this is not the same as pointing out the potential for financial bias that obviously exists in this context. Pfizer reported 86% growth in revenues in this report, but 10% excluding its Covid-19 vaccine. This is unsurprising and perfectly reasonable, at least within the current context of drug development and manufacture.[5] But there is the potential for profit. And so there is the potential for financial bias. CNN operates in the large space between scientific publication, where financial interests must be stated, and the medical professional holding the needle, who might reasonably omit mention of the potential for financial bias. CNN operates in the domain of public discourse and debate. In this context, it reported this information with a positive spin and omitted the possibility of financial bias.

To reiterate, the potential bias does not imply the reality of bias. But communication of the potential for bias is one of the central mechanisms for preventing it from becoming a reality, since it enables users of the information to critically evaluate it.

Safety and effectiveness are subject to stringent regulatory checks, but there can still be misalignment of interests. Safety was subject to fairly consistent scrutiny in the public domain, and our criticisms have much application here. AstraZeneca's vaccine was withdrawn from use in South Africa, the UK, Europe, and Australia in 2021. The operation of financial bias is clear here. The company continued to manufacture it until 2024, when it discontinued it, citing commercial reasons. The company first admitted safety issues in court documents (Investigations Team and Mendick 2024). There was a considerable difference, obviously, in the risk–benefit evaluation of this vaccine from the perspectives of the public health authorities using it and the commercial entity supplying it. Similarly, Moderna was limited in some places, but the manufacturer did not regard the safety concerns as sufficient to cease manufacturing. None of this is particularly surprising; there is no enormous conspiracy being uncovered here. It is just business as usual. Buyers and sellers have different interests. When it comes to the safety of Covid-19 vaccines, this was generally well-communicated in the public sphere, probably due to the intense interest in vaccine safety and a consciousness of the disastrous consequences that might follow from a major safety issue going undetected or without response.

However, effectiveness and necessity may have been subject to financial bias in a more subtle way. In particular, the prolonging of vaccination campaigns and the offering of ongoing boosters are very attractive to pharmaceutical companies, but far less obviously so from a public health perspective. Immunity from infection wanes fast, but immunity against severe disease lasts for a long time. In effect, this means that seasonal Covid-19 could act as a booster, causing mild illness and reinforcing immunity against serious disease. Boosters might reduce the incidence of infection and thus mild illness, but the evidence for an effect on serious illness is minimal, because of the slower waning of protection against serious illness after either infection or previous vaccination. The cost, on the other hand, of a population-wide vaccination campaign is huge. For some people, avoiding infection might be very important, and boosters for those groups might be worthwhile. But for the larger population, it is questionable. Public health is a matter of prioritizing scarce resources, and preventing mild Covid-19 across the general population is simply not a high enough priority to justify that large a slice of the pie. Other uses of that money would save many more lives.

This potential continues to play out in preparing for future pandemics. In some quarters, the emphasis is heavily on vaccine development. For example, the Coalition for Epidemic Preparedness Innovations was established in 2017 with an investment of $460m from governments and private sources (CEPI 2017). This entity specifically focuses on vaccine development, but its name references epidemic preparedness more generally. Their publicity is full of images of large, expensive laboratories.

There is no doubt that vaccination has a lot to offer public health. It is equally clear that there are many non-pharmaceutical interventions that can also make a large difference to public health. In low-income settings, these include the provision of basic medical equipment and supplies, better nutrition, education, and generally measures to reduce poverty. In higher-income settings, too, poverty reduction measures, as well as changes to diet and exercise patterns, have a large potential to reduce avoidable mortality. These measures would certainly have prevented many deaths from Covid-19 before a vaccine was even invented. Vaccination plays a central role in contemporary public health, but it is an ongoing frustration of many involved in public health that more is not done about the social and economic drivers of poor health. Even vaccination programmes depend on socio-economic factors, since a vaccination programme cannot succeed if the vaccine is not distributed effectively, or if the population does not understand it well enough to accept it. But there is money to be made from vaccines. Vaccines' place in public health cannot be denied, but nor can their profitability.

7.12 Conclusion

Any critical commentary on any vaccination programme will provoke a negative reaction from some quarters. This is inevitable in a politically charged context. Arguments and factual claims can be disputed, but we believe that it is important to articulate and criticize the Vaccine Dogma. The harms of the Vaccine Dogma include suboptimal policies. But they also include much wider effects. The CDC's position on Covid-19 vaccination was manifestly unreasonable. Dogmatism of this kind can only antagonize anti-vaccination groups and reduce the credibility of public health institutions. It does nothing to create a truth-conducive scientific environment, given the disconnect between the preponderance of scientific evidence and the policy stance of this scientific authority. This is a harm of scientific orthodoxy. There is a recurrent temptation to dismiss complications in favour of simplicity. But if, in doing so, one dismisses truth in favour of falsehood, or accuracy in favour of error, or sensitivity to changing evidence in favour of unreasoning persistence, then one must resist the temptation.

A final defence of the Vaccine Dogma might contend that it is sometimes necessary to simplify or exaggerate in public health, in order to get a message across, and to bring about changes in behaviour or policy. There is a long-standing ethical discussion surrounding the tension between informing and persuading in public health communication (Blumenthal-Barby 2012; Childress et al. 2002). Public health authorities often face difficult trade-offs between scientific precision and motivational clarity, particularly under conditions of uncertainty or urgency. Some commentators have argued that the use of simplified or even exaggerated claims can be justified if it serves the greater good – such as increasing vaccine uptake or reducing harmful behaviours.

We do not take a position in this broader debate. The idea that public health authorities can simplify or exaggerate in order to change behaviour does not offer an effective defence against the arguments we have mounted in this chapter, even if it is accepted (and it is a contentious idea). It does not deal with many of the points we made. For instance, if mandates are justified on the basis that to refuse vaccination is to harm an innocent bystander, but that basis is false, then this is no exaggeration or simplification, but a downright falsity, resulting in an unjustified use of coercive power. Again, one cannot justify rampant bias in scientific publication as a kind of simplification. Our concern is not solely with the ethics of public health communication, but with the way science itself operates. And our concern is not with simplification or exaggeration at all, but with dogma. Simplifying or amplifying a message is not the same as adhering to it in the face of recalcitrant evidence. Our concern is not with simplification or amplification per se, but with dogma. Public *and scientific* discourse around Covid-19 vaccines ossified into a rigid dogma that resisted critical scrutiny and marginalized legitimate doubt. Whether or not persuasive exaggeration is sometimes ethically permissible, the existence of dogma is not, as this and the preceding three chapters have shown.

Notes

1 If you like, you could say that the CDC are realists about safety and effectiveness, in the sense of "realist" used by philosophers of science to indicate independent existence. In a colloquial sense, however, you might say that their one-sided view of Covid-19 vaccines means they are not realists at all.

2 Maybe not even Peter Singer – *maybe.*

3 The safety concerns around this vaccine are particularly unfortunate for confidence in novel approaches to vaccine development. This vaccine was developed through a new public–private partnership model of drug development. It was also the first vaccine to be announced.

4 Rotashield, developed against the rotavirus, had an estimated risk of causing intussusception of 100 in a million and was withdrawn within a year after its introduction.
5 Pfizer also reported 22% growth in research and development costs, and 286% growth in cost of sales. Drug development is expensive and often fails to produce anything at all.

8

SUPPRESSION OF DISSENT AND THE GREAT BARRINGTON DECLARATION

8.1 Introduction

We turn now to the third element of scientific orthodoxy, suppression of dissent, focusing on the response to the Great Barrington Declaration (GBD). The GBD was a public-facing statement authored by three epidemiologists (Kulldorff, Gupta, and Bhattacharya 2020). It was published and promoted by the American Institute for Economic Research, a non-governmental organization commonly described as "libertarian" or "right wing." The GBD authors also met Scott Atlas, a special advisor on the White House Coronavirus Task Force under the President, to present their view. Atlas became, or perhaps already was, a supporter of the GBD. The GBD attracted the approval of the US President, Donald Trump, whose divisive status deprived the Declaration of any chance of universal acceptance.

As we make clear at the end of the chapter, we do not endorse much of what the GBD says, except that we share its concern that lockdown is unfair to the less well-off. Our interest is in the episode that surrounded it, which, we believe, shows the mechanisms by which scientific orthodoxy suppresses dissent. These are not unique to science, we believe, because we see parallels with the treatment of heresy by the medieval church. We use these parallels, drawing comparisons between the fourteenth-century response to the English heretic John Wycliffe and the twenty-first century response to the GBD. Central to the mechanism of suppressing dissent is the notion of heresy, and we define a notion of scientific heresy and show how dissenters who persisted and refused correction were treated as scientific heretics.

DOI: 10.4324/9781003732495-9

8.2 Definition

We say that there is *suppression of dissent* by a scientific authority when a scientific authority takes active steps to discredit dissenting views and their proponents. This is by far the simplest of our definitions, and we do not regard it as requiring elaboration. What is more interesting is the *way* that dissent arises and is suppressed, which takes us to the notion of heresy.

8.3 Heresy

8.3.1 Ecclesiastical Heresy

"Heresy" can refer to either a heretical doctrine or an ecclesiastical crime. A heretical doctrine is one that fundamentally contradicts the teachings of the Church. For the doctrine to be heretical, the teachings must be fundamental, incontrovertible: in other words, dogma. The crime of heresy is the wilful and persistent endorsement, teaching, or promotion of heretical doctrines, despite being corrected by Church authorities. Thus, to be a heretic is not merely to endorse a doctrine contrary to Church teachings, but to endorse a doctrine fundamentally at odds with Church teachings, and to do so wilfully and persistently, and perhaps also to promote it. Heresy is, thus, distinct from the lesser wrong of erroneous belief.

Although explicit definitions are important, words have overtones determined by their histories. This is especially so when parallels exist between contemporary and historical episodes. The life of Bishop John Wycliffe (1328–1384) and the response to his works are replete with parallels to the activities of and responses to heretical scientists in the Covid-19 Pandemic.

Wycliffe was a fourteenth-century English heretic whose work was highly critical of the Catholic Church, and laid the foundations for the Protestant Reformation a century later (Workman 1926; Hudson 1988; Levy 2002; Evans 2005). He criticized the Papacy and the hierarchical structure of the Church, and argued that it was not founded on scripture; he promoted the authority of scripture over tradition and clerical interpretation; he translated, or initiated the translation of, the Bible from Latin into English, which was forbidden; he argued for the separation of Church and State; he denied transubstantiation, the doctrine that Eucharistic bread and wine became the body and blood of Christ; and he castigated clerics for corruption. His teachings and followers played a role in the Peasant Revolt of 1381.

In 1382, a synod was convened in Blackfriars to deal with his activities and those of his followers. It must have been somewhat chaotic because an earthquake occurred, hence it became known as the Earthquake Synod. Some in attendance felt this was a bad sign. But the authorities were clearly keen not to waste the opportunity, and successfully argued that the upheaval was a sign of the momentous import of what was being done. Twenty-four theses associated with Wycliffe were evaluated; ten were found heretical, and fourteen were found to be merely erroneous. The former included transubstantiation and the authority of the Pope, while the latter included clerical poverty and the temporal power of the Church. This illustrates the distinction between heresy and mere erroneous belief: the former, but not the latter, directly contradicts the teachings of the Church on matters upon which no contradiction is permissible. Wycliffe's activities also illustrate that heresy is not merely thinking something wrong, but doing so wilfully, resisting correction, and promoting it. These distinctions offer several parallels when considering episodes from the Pandemic.

Wycliffe's response to being charged with heresy is also interesting. He responded by appealing, not to the Pope, but to the King of England. This was in accordance with his belief that the Church ought not to possess temporal power. This shows a connection between heresy and power in general, not just in relation to this heresy. Dogma is supported or enforced by political power. Heresy likewise seeks political backing. A heresy is not merely a proposition advanced for debate within existing power structures. If it were, then Wycliffe could have appealed to the Church authorities.

A similar view is taken by the theologian Alister McGrath in his book on heresy, where he writes: "Heresy is a form of Christian belief that, more by accident, ultimately ends up subverting, destabilizing the core of Christian faith" (McGrath 2023, 11). Thus, a heresy is not advanced for debate within existing power structures; it is something fundamental. The picture McGrath paints of the heretic is of an insider who finds themselves, often to their surprise, deeply at odds with their Church. Their persistent, unsuccessful efforts to persuade others to see things as they do ultimately lead them to be labelled heretics: "Heresy is a failed attempt at orthodoxy and an unwillingness to accept it has in fact failed" (McGrath 2023, 31).

Wycliffe's heresies highlight the symmetry between heresy and dogma. They show us that a heresy is not the outright rejection of a framework, but a reinterpretation of its subject matter, so that the fundamental tenets are denied and replaced. Wycliffe was a bishop. He was not an atheist, something almost inconceivable in medieval times. Accusations of atheism were rare, and atheists could expect the most extreme punishments. Wycliffe had a rough ride, but he was not executed, although who knows what might have happened had he lived longer. In any case, his opinions and activities were, and were perceived as, internal threats to the Church. Nor were Wycliffe or his followers persecuted *en masse* in the way of the Jews, nor attacked in anything resembling the Crusades. Heresy is thus not akin to the rejection of an entire framework of belief. Heresy is the reinterpretation of the subject matter of dogma.

Wycliffe died two years after the Earthquake Synod, but his activities continued to inspire his followers and provoke responses from the Church after his death. The Anti-Wycliffe Statute of 1401 explicitly dealt with Wycliffe's activities and identified certain of his teachings and activities as heresies, with a view to suppressing the activities of his followers. Wycliffe himself was officially declared a heretic at the Council of Constance in 1415, which recommended that his bodily remains be removed from consecrated ground. In 1428, the recommendations were confirmed by Pope Martin V. His body was exhumed, burned, and the ashes were drowned in the River Swift.

Particularly concerning for the Church was Wycliffe's promotion of the translation of the Bible into English. In 1408, seven years after the anti-Wycliffe Statute and seven years before the Council of Constance, the Constitutions of Oxford confirmed that it was prohibited to translate or read the Bible in English without the authorization of the Church. They also restricted who could preach and required that sermons be in line with Catholic Church orthodoxy. In addition, they required that all written publications – even those in public collections – be inspected for heresy by Church authorities.

The three condemnations – anti-Wycliffe, Oxford, and Constance – are each separated by seven years, and, although the historical narrative will be much more complicated, one might construe the logical narrative as follows: the Anti-Wycliffe Statute of 1401 was a quick and specific response to Wycliffe – an emergency response, in other words; the Constitutions of Oxford of 1408 then identified the general principles, and extended them in anticipation of unforeseen forms of heresy (it contains considerable redundancy and an emphatic catch-all clause); and the Council of Constance of 1415 retrospectively applied these generalized principles to Wycliffe. Historians and philosophers usually have different intellectual inclinations, and this is the sort of abstraction that historians often detest; but we are not here writing a history, either medieval or modern: we are in the business of constructing sense-making theoretical frameworks. For this purpose, a narrative like this will suffice, and besides, it is remarkable that any parallels at all, however abstracted, can be drawn between the events of the fourteenth and fifteenth centuries and those of the early twenty-first.

8.3.2 Scientific Heresy

We thus offer the following definitions of scientific heresy.

> A scientific heresy, or scientifically heretical doctrine, is a doctrine that contradicts a scientific dogma.
>
> To commit scientific heresy is to wilfully endorse a scientific heresy despite correction by scientific authorities.
>
> A scientific heretic is one who commits scientific heresy.

To commit scientific heresy thus requires two more elements besides believing a heretical doctrine. First, one must resist correction. A pastor in a remote tropical village might believe, for example, that Covid-19 vaccines are in fact intended to cause disease rather than prevent it. But that does not yet make them a heretic, only misguided. If that person responds to a reasonable, charitable, and clear explanation, then they are not a heretic, and never were. If, on the other hand, they reject all reason (as understood by the Orthodoxy), return to their village, and preach against Covid-19 vaccination at every subsequent sermon, they probably become a heretic.

Second, heresy implies wilful endorsement. This means taking some sort of action to promote or support the heresy or undermine the dogma. This could be a public action like preaching, as in the case of the pastor, or it could be a secret action, such as pretending you have been vaccinated when you have not, or it could be some sort of action in the private sphere that is somewhere in between secret and public, such as teaching heresy to one's children.

Thus, committing scientific heresy, in our definition, is more than simply agreeing with a heretical doctrine. This stipulative definition fits actual events of the Pandemic well, as will become clear.

8.4 The IHME Model Was Mere Wrongful Belief

We have already discussed one case of suppression of dissent in Chapter 2, concerning the attack on the IHME's predictive model in April and May of 2020. Here, one can see the parallel between the Church's approach to heresy and the Pandemic Orthodoxy's approach to scientific heresy. The proponents of the IHME were not ultimately deemed heretics: indeed, Deborah Birx and Anthony Fauci remained publicly prominent, in a positive way. But if the proponents of the IHME model had persisted in the face of the "correction" pressed by the Pandemic Orthodoxy, they would have become heretics, and the IHME model, instead of being a wrongful belief – a "dangerous belief," in contemporary parlance – would have become a heresy. What made it not so was the fact that its proponents accepted correction.

At issue was the appropriate methodology for predictive modelling. Therefore, we introduced the episode in the context of methodological rigidity. However, it also bears on the suppression of dissent. The attack on the IHME clearly had the effect of reasserting the epistemic supremacy of a certain view of epidemic modelling, and the custodians of that approach were thereby reaffirmed as the scientific authority internationally, as institutionalized in the arrangement with the WHO.

The methodological differences gave rise to a conflict with the Lockdown Dogma. The IHME's method involved using curves fitted to data from other countries and using them as the starting point for projecting a line in the US states. As US data emerged, the curves were adjusted to improve precision and accuracy. The problem from the perspective of Orthodoxy was that this method implied that curves looked pretty similar across countries, which in turn implied that the introduction of NPIs did not result in dramatic changes. We remarked that this is not quite the same as saying that NPIs had no effect, which is not something that the IHME claimed. However, it shows that dramatic effects are not to be had. This conflicted with the results of ICL9, and it also conflicted with the Lockdown Dogma, which implies that locking down does have a dramatic effect. Otherwise, what would be the point?

It might seem paradoxical to assert that NPIs can have an effect without resulting in step changes in R_t. But there are two good reasons R_t behaves like this. First, the data used to calculate R_t vary considerably on a daily basis. Instead, methods used for calculating R_t are based on data of at least two weeks or more, smoothing out inherent daily variations. But then any change is bound to be smooth, and linking specific policy changes to R_t is virtually impossible. Second, R_t varies throughout the Pandemic, on a curve which starts off high, starting to decline slowly, then falls more sharply as it approaches 1, below which it starts to flatten out. Interventions occurring while R_t is falling fast will appear to have a larger effect than interventions occurring while R_t is falling more slowly, simply because of this background behaviour of R_t. This is especially so if one assumes, as some evaluations did, that R_t is constant between interventions (see Chapter 5). In either case, the effect of the intervention will not be sharp, since R_t is not constant, and not even falling in a straight line. The effect will be to alter the curvature of a curve. Even a sharp change is not easy to detect.

The IHME's model thus represented a potential challenge to the Lockdown Dogma. However, it did not grow into a full-blown heresy. A correction was issued by the scientific authorities, and it was made clear by the accompanying media briefing that this was a correction and not merely a comment. The public health officials who had been smeared as "White House" officials and "Trump administration" did not reply. In her book on the Pandemic, Deborah Birx does not even mention the episode, although she does mention an earlier call in which she consulted ten scientific modelling teams, including ICL and LSHTM. This is despite the fact that she spent a month presenting the IHME's graphs very publicly. There was no persistence, and thus no heresy. It was all a big misunderstanding. The rift was healed.

8.5 The Great Barrington Declaration Was Scientific Heresy

The content of the GBD reduces to two components: a negative claim and a positive one. The negative claim is that lockdowns are a poor public health strategy in response to Covid-19. The positive one is that a strategy called "focused protection" should be used instead. This proposal we call the Focused Protection Heresy.

8.5.1 The GBD's Critique of Lockdown

The negative aspect – the critique of lockdown – is essentially from a public health perspective, the concern being that lockdowns have "devastating effects on short and long-term public health." The devastating effects included "greater excess mortality in years to come, with the working class and younger members of society carrying the heaviest burden," thus invoking the historically central concern of public health not just with the general population, but with the health of groups with lower political, social, and economic status. The timeframe for vaccine development, which at that time was uncertain, is presented in the GBD as being too long to avoid "irreparable harm," again with the emphasis on the "underprivileged."

There are two brief allusions to considerations besides health: a mention that "Keeping students out of school is a grave injustice," and a suggestion in the final sentence that "People who are more at risk may participate" in unrestricted activities "if they wish." Both imply some non-health-based evaluation. But the statement as a whole – which is short, at 509 words – avoids any explicit discussion of the other major discussions of the day, being economic impacts and infringement of liberty. This is certainly deliberate, the idea being to keep the statement within the area of expertise of the three authors, and to avoid the ideological tangle of (especially American) political debate. The GBD seeks not to contrast health with economics or other goods but to present a choice between public health strategies.

8.5.2 Focused Protection

The positive proposal of the GBD was "to allow those who are at minimal risk of death to live their lives normally to build up immunity to the virus through natural infection, while better protecting those who are at highest risk. We call this Focused Protection" (Kulldorff, Gupta, and Bhattacharya 2020). There were two parts to the reasoning. The first concerns costs and benefits, which differ for the two portions of the population designated "vulnerable" and "not vulnerable." For a large portion of the population was at sufficiently low risk ("minimal risk of death") for normal life to resume. The second part of the reasoning concerned herd immunity. Because vaccination was an indefinite time in the future, there was uncertain benefit to postponing infection to that point for the non-vulnerable population at "minimal risk of death." The non-vulnerable population should reach the herd immunity threshold through infection, reducing the burden on them. The benefits of this herd immunity would then accrue to the vulnerable.

8.5.3 Adherence and Propagation

The content of the GBD clearly contradicts the content of the Lockdown Dogma. It became a heresy because it was adhered to in the face of strong "correction" by the Pandemic Orthodoxy, and because it was delivered publicly and politically, rather than within a scientific context. Like Wycliffe, the proponents of the GBD sought to convert people outside as well as inside the epistemic community by direct appeal. And just as Wycliffe appealed to a temporal lord rather than ecclesiastical authorities for support, the GBD appealed to political rather than scientific authorities.

The GBD was not presented as a scientific proposal to the scientific community. It does not pretend to be a laying out of facts for policy-makers to make a decision about; it is a series of recommendations. It was drawn up and announced with the support of a political organization. It was released publicly with no scrutiny or input from the scientific community at large, much as a preprint would be. But unlike a preprint, it was drawn up and announced with the support of a political organization. Its authors took the message directly to Scott Atlas, a member of the US government's Covid-19 response team with a favourable political relationship with the president.

Indisputably, the GBD is a political intervention and was thus propagated in the manner of a heresy. It was also adhered to in the face of a strong reaction from the Orthodoxy, to which we now turn.

8.6 The Pandemic Orthodoxy Reasserts Itself

On 8 October 2020, the director of the NIH in the US, Francis Collins, sent a now-famous email to the Director of the US National Institute of Allergy and Infectious Diseases (NIAID), Anthony Fauci, the Clinical Director of NIAID, Cliff Lane, and the Principal Deputy Director of the NIH in the US, Lawrence Tabak. His email read:

> Hi Tony and Cliff,
>
> See https://gbdeclaration.org/ This proposal from the three fringe epidemiologists who met with the Secretary seems to be getting a lot of attention – and even a co-signature from Nobel Prize winner Michael Leavitt [sic] at Stanford. There needs to be a quick and devastating public take down of its premises. I don't see anything like that on line yet – is it underway?
>
> *Francis (Wenstrup 2024)*

The link was to the GBD, and the proposal was indeed "taken down" quickly and devastatingly, in an episode that is the most well-known and perhaps still the most divisive of the era. The takedown involved fast responses by non-official voices on social media and online media (blogs), followed by responses in more established media by scientific authorities, and then followed on 14 October by a formal publication (Alwan et al. 2020) matched with an online version called the John Snow Memorandum (JSM 2021). We see parallels between these episodes and those following Wycliffe's heretical activities, leading up to the Anti-Wycliffe Statute of 1401. The parallels help us see, with the coolness of history, how the Pandemic Orthodoxy was challenged and reasserted itself.

The GBD did more than state a scientific position, however. It sought to displace the Pandemic Orthodoxy, specifically, the Lockdown Dogma, which at the time was the primary focus of everybody's attention. The GBD proposed a doctrine contradicting the Lockdown Dogma: a scientific heresy, in the sense of a scientifically heretical dogma. Furthermore, its proponents wilfully endorsed and promoted this doctrine, and persisted despite corrective efforts by the Pandemic Orthodoxy. It quickly became apparent that there would be no backing down. The GBD remains live at the time of writing, and its authors stand by it, to the best of our knowledge. Thus, the GBD was a scientific heresy, and its proponents committed the corresponding scientific "crime."

This explains features of the response that would otherwise be difficult to explain. Francis Collins later expressed his regret at the way he expressed himself in that email, but he pointed out, quite rightly, that the GBD was never presented as a matter for scientific debate, but was interjected into the political domain as a direct attempt to influence policy (Braver Angels 2023). This was not a purely factual disagreement between two groups of scientists. It was a fight for power.

8.7 The Pandemic Orthodoxy's Response

8.7.1 The John Snow Memorandum

Those fast online takedowns were not centrally coordinated. The first organized response of the Pandemic Orthodoxy was the John Snow Memorandum (JSM), whose name invoked a legendary historical epidemiologist (but may have been more widely reminiscent of an inspiring fictional warrior). It was titled "*Scientific consensus on the COVID-19 pandemic: we need to act now*," and it was published online and in *The Lancet.*

This consensus claim was not fully accurate. For one thing, the GBD did attract some scientific support. The GBD may not have attracted full support on a large scale, as the list of signatories shows. Many would have treated it as a call to re-evaluate lockdown strategy, rather than a detailed public health proposal. In the same way, many supporters of the JSM would have taken its critical components more seriously than its positive proposal. We know this because the positive proposal was certainly not an object of consensus.

The major non-consensual inclusion was the vaccine-plus strategy (VPS). The VPS was, in effect, the idea that restrictions should persist after vaccination, in contrast to the general public expectation that with widespread vaccination, the Pandemic and its manifestations would be over. It was rarely enacted and then not for long, because public and political support for continued regulations had waned, and because, ultimately, the Pandemic Orthodoxy lost its hold on power and on the voice of science. Public health more broadly did not see the threat of Covid-19 post-vaccination as serious enough to support the VPS, and, in the end, this meant the VPS's scientific authority dwindled. For present purposes, this shows that the JSM had the power to claim to represent scientific consensus when it did not.

The Earthquake Synod likewise included elements of discord, which were not reflected in the resulting Anti-Wycliffe Statute. The earthquake was taken by some as a negative sign. This view was argued down, and the doubts were not reflected in the document. This is also an indication of haste, since, if there had been more time, it would presumably have been feasible to reconvene, and less important to deal with the doubts there and then.

Like the Anti-Wycliffe Statue, the JSM was quick and dirty. This is evident in the disconnected mix of arguments it contains, some better and some worse, leading to occasional inconsistencies. For example, the JSM argues that infection does not confer immunity against further infection. This is correct. However, this supports the idea that we need to learn to live with Covid-19, since otherwise the implication is that lockdown must be indefinite.

Both the Anti-Wycliffe Statute and the JSM were drawn up in haste, and both were clearly directed at a particular target. Both were heavily contextual, being bound up in the events of the time. The JSM included the phrase "with winter approaching," positioning itself both in time and space – in the northern hemisphere. It focuses on attacking the positive proposal of the GBD, Focused Protection, instead of dealing with the GBD's attack on lockdowns. The GBD's positive proposal is weaker than its criticism, as we explain at the end of the chapter. But of course, the criticism and the proposal are logically separable, and one might accept the GBD's criticism but make a different proposal. The positive proposal of the JSM is a simple reference to the VPS, which, as already discussed, was not a consensus view.

Neither the Anti-Wycliffe Statute nor the JSM was strong on general principles or positive proposals. This was, in both cases, a weakness that needed fixing if the respective heresies were to be properly suppressed. They were thus both followed by more considered pronouncements: respectively, the Constitutions of Oxford of 1408 and a formally constructed consensus statement in 2022.

8.7.2 The 2022 Delphi Consensus Panel

In November 2022, *Nature* published the result of a Delphi consensus process applied to the Covid-19 Pandemic "to recommend specific actions to end this persistent global threat to public health," titled "*A Multinational Delphi Consensus* *to End the COVID-19 Public Health Threat*" (Lazarus et al. 2022) (hereafter, Delphi22). It involved 386 "experts" from 112 countries, and resulted in 41 statements and 57 recommendations in six domains: communication, health systems, vaccination, prevention, treatment and care, and inequities. The top three ranked recommendations are, in order, that a whole-of-society approach should be taken in pandemic preparedness and response, that public health communication should improve, and that prevention should, in "all countries," take the form of a "vaccine plus" approach.

This statement displays some interesting parallels to the Constitutions of Oxford of 1408. The two documents perform the same function. Unlike their predecessors, they are not specifically targeted to take down a certain heresy, but at preventing the proliferation of heresy, and guiding the faithful back to proper dogma. Their stance is thus forward-looking. However, we first discuss Delphi22 in its own right, before turning to the parallels in the next section.

8.7.3 The Delphi Method

A Delphi consensus statement is developed according to a specific methodology, which we will describe shortly, with the goal of indicating the (presumably high) level of agreement between participants on an issue as it is expressed in a specific statement.

The Delphi methodology was originally developed for forecasting, the first use probably being in a 1950s US project to predict enemy attacks (Nasa, Jain, and Juneja 2021). It was then widely applied in business, finance, economics, and development planning – areas concerned with what we might call forced prediction under uncertainty: that is, predictions that have to be made because of the necessity of making decisions, despite inadequate information.

The Delphi method does not have a formal statement or evaluation framework, and variations exist, but its core components are as follows:

1 Panel selected.
2 Initial questionnaires distributed.
3 Analysis, with feedback provided to the panel.
4 Subsequent rounds: questionnaires are revised based on the analysis, and re-circulated, the process iterating until a stable (presumably high) level of consensus is reached (or stable levels of consensus regarding each of multiple points).
5 Final consensus statement agreed and published.

Delphi is not a fact-finding exercise, but a statement expressing the beliefs of its participants. These may be modified during the Delphi process, but that process does not involve new fact-finding. Delphi consensus may involve both statements and recommendations. The method originates as a decision tool. In principle, participants may even remain in disagreement on certain facts while coming to a consensus on what to do.

Delphi is not a tool for uncovering existing consensus, nor for informing the public of it. This means a Delphi consensus statement does not represent the result of "convergent" reasoning, in the sense philosophers of science sometimes talk about, because it is not the result of people independently arriving at the same conclusion. Rather, they arrive at the same conclusion via a process of *collective* reasoning.

Delphi takes elements from two different approaches to problem solving: expert-led and collective. Expert-led problem solving occurs when a group is selected with experts representing each domain in which relevant facts and skills might be found. This approach seeks to access the knowledge and skills of the experts. Collective problem-solving, on the other hand, seeks to access collective intelligence or problem-solving, and occurs when people seek to solve problems as a group. Collectives of non-experts can sometimes outperform experts, but this approach remains somewhat counter-intuitive. And there must be *some* inclusion criteria in a group assembled for collective problem-solving, so as to exclude labradoodles and rocks. The Delphi process might be thought of as an effort to get the best of both approaches. While its details differ in different domains, core features are common: "anonymity, iteration, controlled feedback, and group response (or consensus)" (Nasa, Jain, and Juneja 2021). These features, and indeed the whole idea of seeking consensus, reduce the influence of recognition as an expert. This is contrary to the expert-led process. Once a panellist has crossed a certain expertise threshold necessary to join the panel, the playing field is levelled, in theory.

8.7.4 Panel Stacking in Delphi22

There are, of course, challenges for the Delphi process. One criticism of Delphi22 concerns the lack of neutrality of its participants. A paper by 67 authors, including a number of well-known epidemiologists, presents evidence that 35% of the core group of 40 leading the Delphi22 had advocated for zero-Covid, and contends that advocacy positions can be a source of bias (Kepp et al. 2024). *Nature* desk rejected this paper (i.e. rejected it without sending for peer review). *Nature* receives many submissions and rejects many. Still, this was a direct response to a *Nature* article, relevant to *Nature* readers (therefore), and was eventually published in the *Journal of Clinical Epidemiology.*

The Kepp et al. paper criticizing Delphi22 adopts a very specific perspective on the validity of evidence and, continuing a pattern we have already noted with the GBD and JSM, contains positive statements that are not widely accepted outside the community from which its authors are drawn. For example:

> Empirical data suggest that consensus-based approaches without evidence synthesis are 3–5-fold more likely than evidence-based approaches to yield misleading advice.
>
> *(Kepp et al. 2024)*

This obviously depends on what you think amounts to "empirical data," and, if you think that evidence-based approaches are the surpassing source, then you are effectively assuming what you set out to prove. It also ignores the decision-making origins of the Delphi process, the benefits of collective problem-solving, and the risk of over-generalizing from those "consensus-based approaches without evidence synthesis" reviewed, to every Delphi process.

All the same, the critical component of the paper is compelling. A large number of previous advocacy activities are identified in both the core group and the wider panellists. Most are undeclared: "all members of WHN, OzSAGE, End Coronavirus, ZeroCOVID-US, NOCOVID Europe, and all but two of seven active members of Independent SAGE declared no competing interests," seemingly contrary to journal requirements (Kepp et al. 2024). Delphi22 describes selection "primarily based on publication record and engagement on Covid-19 issues as well as online biographies." Indeed, this is not a very convincing method. While the statement does not say that they used Twitter to find people, it also does not rule out the possibility that algorithms involved in online tools, including social media and common search engines, might have created a bias in favour of people agreeing with the core group. The accusers also point to the use of snowball sampling, which, as they point out, is "highly sensitive to personal network bias and may sometimes reflect limited merit" (Kepp et al. 2024). The point is made that even 386 scientists are a drop in the ocean of contributors to the enormous scientific literature on Covid-19 (which they say included 720,801 authors), and that the majority of scientists writing on Covid-19 – at any rate, far less than 35% – were almost certainly not members of advocacy groups. Nor is it plausible that this proportion of scientists writing on Covid-19 supported zero-Covid, which was always a minority view and certainly was in November 2022, well after the widespread introduction of vaccines and the arrival of Omicron.

The paper provides good evidence that the consensus statement cannot be treated as reflective of scientific opinion at large. On the other hand, it does not engage with the substance of the paper at all. If Delphi is taken as a method of representing scientific opinion, it is arguably misused, but if it is a method by which its participants can harness their collective reasoning powers to solve a problem, it is less clear whether panel stacking matters. That is because it is not clear whether an unrepresentative group is worse at solving problems than a representative one. However, we have already pointed out that the Delphi22 was open to being misinterpreted as a consensus between independent views. To this extent at least, there is a case for Delphi22 to answer.

8.7.5 Does Delphi22 Represent Scientific Consensus?

The critique just discussed does not positively engage with the content of Delphi22. We now turn to consider whether it did indeed represent a wider scientific consensus. Once again, we find a mixed answer. This is important because Delphi22 presents itself as a *broad* consensus not only as regards support but as regards the breadth of claims on which there is consensus, and especially on the way forward – "to End the Covid-19 Pandemic." Actually, the central claims about how to achieve this end were not the subject of wider consensus at all.

In some respects, the Delphi22 represents a shift from the Pandemic Orthodoxy as it was in 2020. In particular, there seems to be an appreciation of the problems with methodological rigidity and with the incorrect attribution of exclusive epistemic access to "mathematical modelling." A high level of consensus was achieved in relation to "the risk of over-reliance on experts from a small number of disciplines (table 3 (STMT6.8)), often excluding the expertise of community members (table 4 (REC1.2)) and vulnerable groups (table 3 (STMT6.7))" (Lazarus et al. 2022, 341). The phrase "Whole of society" is defined as including "multiple disciplines, sectors and actors (e.g. business, civil society, engineering, faith communities, mathematical modelling, military, media and psychology)" (Lazarus et al. 2022, 337 & 341). Similarly, the statement shows more nuance in relation to lockdown than is present in those papers we criticized in Chapter 5. Delphi22 recognizes the importance of an approach that balances risks and harms.

However, despite these adjustments, Delphi22 remains part of the Pandemic Orthodoxy. It remains committed to the necessity and effectiveness of lockdowns in periods of high community transmission, which is the core of the Lockdown Dogma, and it remains committed to the use of masks as part of the "plus" in its vaccine-plus approach. This indicates that the core of both these dogmas has been insulated from the evidence, which, as we have argued at length, would render an unequivocal statement inappropriate on either topic. Indeed, even attempting a consensus statement on a complicated evidential picture lends itself to dogmatism.

Delphi22 adheres to the Vaccine Dogma, too. It presses the necessity of vaccinating all populations, including children, and the continued administration of boosters, to "end" the Covid-19 Pandemic. For instance, the UK was recommending vaccinations for people aged 40 and above. Again, opinions about vaccination policy were divided at the time. Scientists saw the risks and benefits differently, and had different ideas about which groups should be vaccinated, and whether a long-term booster programme was superior to regular cycles of seasonal infection of the kind we live with for influenza. It is very unlikely that the degree of consensus present among Delphi22 contributors represents scientists more widely. In any case, even seeking consensus on a complicated question of this kind is crude. Delphi22 clearly strives for a reasonable, inclusive feel. But it does not compromise on the Pandemic Dogmas we have identified.

8.8 Delphi22 and the Constitutions of Oxford

The Constitutions of Oxford begin with a mighty assertion of authority:

> Thomas, by divine permission archbishop of Canterbury, primate of all England, and legate of the apostolical see, to all and singular our venerable fellow-bishops, brethren and suffragans, abbots and priors, deans of cathedral churches, archdeacons, provosts and canons, rectors, vicars and chaplains of parish churches, all clerics and laymen whatsoever within our province of Canterbury, health, and firm adherence to the doctrine of holy mother Church.
>
> *(Arundel 1408)*

Compare:

> we convened, as part of this Delphi study, a diverse, multidisciplinary panel of 386 academic, health, non-governmental organization, government and other experts in COVID-19 response from 112 countries and territories to recommend specific actions to end this persistent global threat to public health. . . . The findings of the study, which have been further endorsed by 184 organizations globally, include points of unanimous agreement, as well as six recommendations with >5% disagreement, that provide health and social policy actions to address inadequacies in the pandemic response and help to bring this public health threat to an end.
>
> *(Lazarus et al. 2022, 332)*

The function of both statements is to show that the author speaks not only with the highest authority but also for all right-thinking people. Neither actually says "all," but both provide extensive lists, and of course, Delphi22 emphasizes consensus.

Of course, one does not want to overstate the parallels between these two documents. Yet it is possible to identify a parallel between many of the 13 articles of the Constitution, and some aspect of the 2022 Consensus Statement.

Central themes emerge from both documents: that information must come from authorized sources only; that authorities must actively prevent what would now be called misinformation, and was then called heresy; that authorities should use their power to punish purveyors of heresy, or misinformation; that individuals are unable to interpret primary sources of knowledge for themselves; and, of course, that there is universal agreement within the epistemic authority about all important matters. This is extremely surprising from one perspective: the documents are entirely different in authorship, historical context, and intellectual setting. And they are very different in tone. One, but not the other, makes moral judgements and prescribes processes for trial and punishment. Yet, both were written by human beings seeking to assert their view of a certain matter.

The common theme is the prevention of the dissemination of falsity – heresy or misinformation, depending on the era. Mis- and disinformation are, in their function, the modern equivalent of heresy, and the Pandemic played a significant role in shaping the concepts. Mis/disinformation emerged in 2016 during the first election of Trump as US President. Before that, the concept hardly existed. In 2020, the terms enjoyed a new surge of popularity: Google Trends show a step increase in searches for the term *misinformation* around March 2020 (Google Trends 2024), and the top related search terms include Covid-19. In 2024, the World Economic Forum identified misinformation and disinformation (combined) as the biggest global risk (WEF 2024) during the Pandemic. These terms are, in effect, modern words for heresy, and mis/disinformation were not only common in the Pandemic, but were, in part, constructed by it.

A further parallel between the Constitutions and the Statement is the context of widening access to information that was previously accessible only through the epistemic authorities. Wycliffe's interpretive and translation work parallels the activities of scientists and commentators who presented their own, non-orthodox interpretations of scientific information. The scientific heretics of the Pandemic, such as Michael Levitt, John Ioannidis, and Sunetra Gupta, presented their own interpretations of the evidence, rather than that of the established institutions or official public health bodies. The activity of doing so was in neither case heretical, but it was nonetheless prohibited, and the resulting claims typically were heretical, as we have seen.

The internet renders scientific research far more widely available than it used to be, and any member of the public can read and try to make sense of any scientific paper without having to be a member of a university with a well-stocked library. A scientist who has been unable to convince their academic peers may still find a receptive audience in the public sphere through preprints, blogs, online videos, social media, etc. This is especially so if people in the public sphere also feel unfairly dismissed by a scientific orthodoxy. Scientific conflicts may thus spill over into the public sphere. Alternative "churches" may be established, become popular, and attract financial support, all without the blessing of the epistemic authorities. In this sense, the modern era is akin to the situation after the advent of the printing press, which postdated Wycliffe by a century and is often credited with igniting the Reformation. Other social and political trends arguably parallel those of the Reformation too: anger at moral decay, and a desire to tear down the trappings of unjustified authority. Whether that parallel plays out either in relation to scientific knowledge or the wider world remains, at the time of writing, to be seen.

8.9 Was the GBD Right?

While we obviously share a critical attitude towards lockdowns, we do not want to be mistaken for endorsing the GBD. In this section, we provide our reasons for not doing so.

We already mentioned that, in its negative critique, the GBD focuses almost entirely on health. We regard this as a mistake, both in the context of the debate at the time, and for public health strategy. Public health must not confine its attention to health outcomes, prioritizing health above all other aspects of human flourishing. In Chapter 10, we will show how many other ways lockdown was harmful, besides health. The argument that lockdown was more harmful than Covid-19 from a health perspective does not convince many people, and, given the history of some of the authors underestimating the infection fatality rate of Covid-19 and over-estimating its prevalence, this serves to seriously undermine the credibility of the health-based critique. Where the GBD points out the burdens that lockdown places on the less well-off, we are in full support; but where it links these strongly or primarily to health outcomes, we find ourselves frustrated that a wider view was not taken of the burdens of lockdown.

On the positive proposal, we share a number of concerns raised in the critiques already discussed. In particular, as we showed in Chapter 4, population-level suppression is effective for reducing eventual Covid-19 mortality. It is simply not the case that letting Covid-19 run through the “non-vulnerable” population without any population-wide control measures would have resulted in the same eventual mortality. There are a number of reasons for this, and, while the empirical evidence of the eventual outcomes was obviously not available at the time, it was predictable from contemporaneous knowledge about Covid-19 and established epidemiological principles.

For example, it was clear that there was no sharp distinction between vulnerable and non-vulnerable groups. Regarding age, in particular, there is a large “grey zone.” Is an overweight 50-something male in the vulnerable group? There are many such people in the world, and many of them are economically productive – and obese. In the UK, 33% of men aged 55–64 are obese, and 30% of men aged 45–64. If they are counted as vulnerable and thus quarantined, normal life would be majorly affected due to their economic activity; if not, mortality would be significantly affected, as it was in countries that achieved lower levels of population-level suppression.

Relatedly, it is an established principle of public health that protecting a large enough number of people at moderate risk will often reduce mortality more than focusing on a small number of people at high risk (Rose 1992). Focused protection contradicts a common, if not overriding, principle of prevention.

Another established epidemiological point is that the rate of spread matters, and a slower rate of spread is a legitimate target of intervention in its own right, even without any endpoint in sight (Morabia 2004). Delaying the spread of a disease means that, for any given point in time, cumulative mortality will be lower. Even conceding that "eventually" everyone will be infected, "eventually" is not an absolute time. People are always dying, and the slower the spread, the more of them die without being infected. This is especially relevant when the most vulnerable population is the elderly, who die soonest. Thus, for any given future time we care to pick, regardless of the availability of a cure or vaccine, cumulative mortality will be lower in proportion to how much we have "slowed the spread."

Protection itself is also affected by population-level suppression, because protection becomes more difficult when prevalence in the general population is higher. So even the Focused Protection Strategy is, to some extent, tied up with the extent to which suppression is achieved in the wider population.

In short, we do not see the GBD as presenting either a comprehensive critique of lockdown or a viable public health strategy. Nonetheless, we sympathize with the effort to prompt a re-evaluation of the lockdown strategy. We just wish its content had been different.

8.10 Conclusion

In retrospect, the GBD was a missed opportunity on both sides for an important discussion of lockdown. It contained two elements: the rejection of lockdowns and the proposal of focused protection. But it wrongly connected them, as if the arguments against lockdowns were also arguments in favour of Focused Protection. This made it easy for opponents to defend lockdowns, simply by defending the strategy of reducing community transmission. It is much easier to argue in favour of controlling community transmission than it is to argue specifically in favour of lockdowns. Hindsight thus reveals a strategic error in the GBD's dialectical positioning.

Had the authors instead recommended *partial* community-wide restrictions, short of lockdown, combined with stronger measures for vulnerable groups, then the proposal would have had more plausibility. Had they proposed a context-specific evaluation of these restrictions, it might have garnered the support of the WHO, whose senior figures were by then urging countries to hold back from using lockdowns as primary control methods:

> We really do appeal to all world leaders: stop using lockdown as your primary control method.
>
> *(Nabarro 2020)*

It would also have opened up a space for nuance. Cost-benefit is mentioned in the GBD, but the proposal itself does not open up much space for this, since it jumps from one extreme to another.

However, the GBD cannot reasonably claim to have been trying to open up a space for debate, which is how some of its authors have sometimes presented it (Grove 2020; Gupta 2020). There is no suggestion in the GBD that there should be a debate. There is a suggestion that the policy should change. There was no invitation to debate in the process by which the GBD was published. It was not submitted to a journal but placed on a website, and it was accompanied by invitations to sign, not to comment: a point Francis Collins, who wrote that hasty email, has made in subsequent reflections (Braver Angels 2023). If the GBD had won the day, it is hard to imagine that its proponents would still be inviting proponents of lockdown to debate.

We have already explained our view that population-wide suppression measures were important during the Pandemic. The sad thing about the GBD episode is that it reinforced the idea that population-wide measures are an all-or-nothing affair. This enabled the critics to completely dodge serious discussion of how to weigh benefits against the harms of lost schooling or the burdens on poor families and small businesses. The harms could continue to be acknowledged as regrettable inevitabilities. Lockdown was justified by a threshold logic, which rendered it unnecessary and even dangerous to weigh the marginal benefit of a given intervention against its marginal cost. Had the authors of the GBD launched a proposal urging more moderate population-wide restrictions, with each component's marginal cost-benefit ratio analysed, and had they proposed a set of no-brainer measures (restrictions on large gatherings, for instance, which even Sweden implemented) and some to which they were strongly opposed (school closures, for instance), then the GBD might have pried open a space for public debate before the next series of lockdowns came in.

9

INDIRECT POLITICAL AUTHORITY

Following the Science

9.1 Introduction

"Follow the science" became a common phrase in the first half of 2020. It was both unclear and controversial: unclear as to exactly whose directions were to be heeded, and controversial because it placed science in the role of leader. No theory of political legitimacy places science in that role. Still less does science itself, being based on a disciplined rejection of authority, at least in the eyes of many scientists. What makes science special is that it does *not* follow any leader but seeks out answers for itself. Acts of faith and declarations of allegiance are paradigmatically unscientific.

Yet it is obvious why the slogan emerged. Scientists were predicting large numbers of deaths from a new disease and urging governments to act to reduce those numbers. On the assumption they knew what they were talking about, it might seem a matter of common sense and common morality to listen to the scientists. Objecting that scientists have no right to make recommendations and should simply lay out policy options would appear to be mere quibbling if the numbers were big enough, and they were very big indeed. The slogan "Follow the science" was meant to encapsulate all this, end talk, and produce action. It was not meant as a contribution to political philosophy.

In this chapter, we propose a model for the governance of scientific advice, especially during emergencies. We start with a fictional scenario in which it is intuitively compelling that the science should be followed: the premise of the movie *Jaws.* We go on to show how existing accounts of the role of science in society fail to validate this intuitive reaction shared by generations of audiences. Their proposals fail because they depend on a kind of deliberation that is simply

DOI: 10.4324/9781003732495-10

inappropriate in the presence of a large shark. We go on to consider a set of proposals by Jonathan Birch for empowering scientists to be more directive in emergencies than they would in normal times. However, Birch's analysis stops short of providing principles of governance. It rescues the directive nature of some scientific advice during the Pandemic, especially ICL9. However, we take a different view of ICL9, and of the performance of SAGE's models and predictions more generally. We agree with Birch that directives of the kind contained in ICL9 are not wrong in principle, but we believe that they were wrong in practice. We go on to explain why, proposing our own analysis of the situation in which scientists may direct policy, which we call *indirect political authority* (IPA). We show how it applies in the *Jaws* case, but not for Pandemic-era epidemiological modellers. We conclude with some reflections on how the IPA framework might have been implemented to improve scientific advice during the Pandemic.

9.2 Jaws

A night-swimmer dies in the ocean while her boyfriend is unconscious on the beach. Her remains are recovered. A marine biologist, Matt Hooper, is convinced that a huge shark killed her. He urges the authorities to close the beach. The authorities decline, contending that she was killed by a boat that did not see her in the darkness. They are concerned about the potential economic impact of closing the beach during peak season. The scientist presents the authorities with the facts and his inferences. But he does not stop there. He makes strenuous efforts to convince the authorities to lock the beach down. Does he do wrong?

This scenario is the premise of the movie *Jaws.* As an audience, we are invited to think that Hooper does nothing wrong, while the Mayor and Police Chief are wrong to ignore him. The tension of the movie is created, not merely by there being a monster, but by the frustration the audience feels – with Hooper – at the authorities' refusal to listen to him. Yet Hooper possesses no political authority of any kind. He has not gone through any legitimating process, whether election or due appointment process. He holds no public office. He is a scientist, and his training as a scientist confers no political authority upon him – at least not in any direct way.

Not only does he lack authority to *direct* or *control* policy, but he also lacks any special right to *advise* or *lobby* in his capacity as a scientist. His expertise gives him the right to lay out facts, but he has no more right than anyone else to express an opinion about what decisions ought to be taken on the basis of those facts. If he expresses an opinion, then it is as a concerned citizen, since Hooper holds no political office. And if he uses his special access, granted as a scientist, to influence decisions – or worse, uses the social and political status that his scientific position confers upon him – then, on the standard view, he oversteps.

It follows, on this seemingly commonsensical view of the situation, that Hooper ought not to put pressure on the Mayor, and that he does wrong in seeking to influence the Mayor's decision. For he is not merely providing the Mayor with the benefit of his expertise, and letting the Mayor do what he was duly elected or appointed to do. He is seeking to influence the Mayor. At best, his actions are excusable as strenuous lobbying. But they certainly do not create a duty on the Mayor to do what Hooper says.

However, in common with audiences of *Jaws*, we feel that this is counter-intuitive. Surely Hooper is right to urge the Mayor, and the Mayor ought to grant Hooper a kind of authority in this situation: not in virtue of any legitimating process, but in virtue of what he knows. Surely the Mayor does not have unfettered discretion to consider and reject Hooper's advice in this situation. Surely he is bound to listen. But if that is so, then Hooper has a kind of political authority, albeit transient. What is this authority, and what principles create it and constrain it?

9.3 Existing Theoretical Frameworks

Existing accounts of the relationship between science and policy, whatever their value in other contexts, offer limited guidance in the kinds of urgent decision-making contexts we consider here, such as pandemics or public safety crises. Two broad limitations account for this.

First, most such accounts focus on non-emergency settings. Even when urgency is acknowledged, the proposed mechanisms typically pre-suppose time for reflection, deliberation, or institutional negotiation. A situation involving a large shark or a lethal pandemic does not afford such luxuries. What is needed in such cases is a way of identifying conditions under which action should be taken without deliberation, and a justification for when scientific advice may justifiably direct or constrain political decisions. Existing frameworks generally explore the appropriate powers of scientists under normal conditions, but do not address the question of when and how those powers should change in emergencies.

Second, existing accounts tend to assume functioning democratic systems, either explicitly or implicitly. Yet many governments are not fully democratic, and even in mature democracies, the institutions of representation and deliberation often fall short of their ideals. The challenge of integrating scientific expertise into decision-making arises across all forms of government, because in none of them does the process that confers political legitimacy also confer scientific expertise (Plato's proposal for philosopher kings not having proved popular). Moreover, emergencies often require coordination across political systems with very different structures and legitimacy frameworks. This is a further reason to decouple the question of when to empower scientific voices from the question of how that empowerment fits into a democratic theory of legitimacy.

Several influential frameworks illustrate these limitations. While each offers important insights into the ethical or democratic dimensions of scientific practice, none is designed to address the specific challenge of scientific authority in urgent, high-stakes contexts. We consider three prominent examples here, not to dispute their value, but to clarify why they do not resolve the particular problem we aim to address.

9.3.1 Kitcher's Well-Ordered Science

The first is Philip Kitcher's account of "well-ordered science," which focuses on the alignment of research priorities with democratic values. Kitcher proposes that science should be "well-ordered," meaning that its research priorities should be set not unilaterally by scientists, but in accordance with values arrived at through inclusive and informed public deliberation (Kitcher 2001). Kitcher imagines a model in which randomly selected citizens are tutored by scientific experts before deliberating together to determine what kinds of research best serve the public good. This is not meant as a literal proposal for institutional design, but rather as a normative ideal: a vision of what scientific priorities ought to reflect, even if they are not in fact determined through such a process.

This is an appealing vision for aligning science with democratic values over the long term. But it is not designed to address the kind of urgent, high-stakes decision-making contexts that arise during emergencies such as pandemics. In such situations, there is no time even to approximate the outcome of an ideal deliberation among informed citizens. One might see Kitcher's model as articulating precisely what cannot be done in such situations, and thus posing rather than answering the question we are seeking to answer here. Moreover, the model itself provides no criteria for determining when scientists may permissibly issue directive recommendations, or when political leaders ought to follow them. A characteristic of emergencies is the lack of opportunity to reflect.

The challenge we are addressing in this chapter is precisely that: how to assess the legitimacy of directive scientific advice under emergency conditions. Kitcher's account helps us see the importance of democratic legitimacy in science, but it does not offer a mechanism for emergency epistemic authority. In this respect, it underscores the need for a framework like IPA, rather than providing one itself.

9.3.2 Douglas's Values in Science

Let us now turn to Heather Douglas's seminal work on the role of values in science. Douglas's central claim is that non-epistemic values are an inescapable part of scientific reasoning, especially in contexts of uncertainty and risk. In particular, she argues that scientists have a responsibility to consider the potential consequences of error – both false positives and false negatives – and that such judgements necessarily involve social and ethical values. Douglas contends that, far from being a threat to scientific integrity, the inclusion of values is necessary for responsible science, and scientists ought to be transparent about the value judgements that influence their work (Douglas 2009).

This approach has been highly influential and aligns with a broader literature that emphasises pluralism and the social embeddedness of science (e.g. Longino 1990; Elliott and McKaughan 2014). It offers a powerful challenge to the idea that science is or should be value-free, and it advocates the idea of an ethical framework for how scientists ought to reason and communicate under conditions of uncertainty.

However, while Douglas's account is highly relevant to the ethics of scientific practice, it does not address the question we are pursuing here: under what conditions, if any, should scientific advice be treated as politically authoritative in emergency situations. Douglas gives scientists moral duties – to reflect, disclose, and reason well – but these duties do not confer political rights. Transparency about values is not the same as a warrant to influence policy directly, especially when doing so may restrict the liberties of others.

In the *Jaws* scenario, for instance, Douglas might say that the Mayor placed too much weight on the cost of a false positive – closing the beach unnecessarily – and too little on the cost of a false negative – failing to act on a real threat. In her view, Hooper's responsibility would be to consider these trade-offs carefully and to be transparent about them. But such responsibilities do not explain why Hooper is entitled to *push* for a particular policy, nor why the Mayor is *obliged* to listen. Our view is that Hooper was entitled, and the Mayor was obliged, or at least, we want to ask whether there are situations where this might be the case, and we want to provide an account of them, if there are. But mere ethical sensitivity does not constitute political legitimacy.

Nor would a more participatory approach solve the problem. As Pamuk observes, asking scientists to represent public values – whether individually or through "minipublics" – risks collapsing the distinction between scientific expertise and political authority. Scientists are not elected representatives, and are not authorized to speak on behalf of the public, however ethically motivated they may be:

> Expecting scientists to discern and use social and political values in their research would be to assign scientists a duty of political representation. This is a role for which they are neither qualified nor properly authorized.
>
> *(Pamuk 2021, 54)*

The issue, then, is not whether values matter in science, but whether value-awareness alone can justify directive scientific advice in urgent political contexts. Our view is that it cannot. Scientists are not entitled to make political judgements in the name of others, even if they are transparent about the values they bring to bear. That task properly belongs to political representatives.

Pamuk's proposal for "science courts," in which scientists present competing views to citizen-juries, offers one possible way of reconciling expertise with democratic legitimacy. But even this approach, which rests on a stronger political theoretical foundation than Douglas's, runs into problems under emergency conditions. Science courts require time, institutional infrastructure, and public deliberation, none of which are readily available during crises like pandemics. In such cases, executive decision-making is inevitable, and what is needed is a framework for guiding and constraining the authority of scientific advice in the absence of deliberation.

Douglas's work, like Kitcher's, offers a valuable perspective on the ethical dimensions of science and its relation to democratic values. But it does not provide an answer to the specific question this chapter addresses: when and how science should be followed in emergencies. That, we suggest, requires a different kind of framework.

9.3.3 Birch's Normatively Heavy Recommendations

Finally, Jonathan Birch has proposed an account of when scientists may make "normatively heavy recommendations." His account is based on a detailed reading of the minutes of UK SAGE in the early part of the Pandemic. He argues that:

> Scientific advisers may permissibly make normatively heavy recommendations, including simple, unambiguous imperatives to implement a specific policy option, where they have a justified belief that this option is the only robust way to avert a catastrophic outcome.
>
> *(Birch 2021)*

This is eminently plausible, but it falls short of what we are after, which is governance principles that might at once give rise to and constrain political authority. Birch's proposal creates a right to speak, but no duty on any listener. It renders scientific recommendations *permissible*, but it does not give them any authority. Scientists are permitted to speak freely in emergencies, but leaders are not obliged to listen. This leaves open what the attitude of leaders should be to scientists who are permitted to be unusually directive. Birch's account is not inconsistent here: it simply leaves the question open, and this, we believe, limits its utility in an emergency. It provides a licence for scientists to shout, but there is no guarantee anyone will listen.

The flip side of this point is that Birch does not focus on constraining scientific power as much as we would like. He provides standards for scientists to apply when reflecting on whether they may speak, but no standards for others to judge them by. In the early Pandemic, modellers probably satisfied Birch's criterion, as he himself points out. Birch regards ICL9 and SAGE modelling more generally as justified. We, on the other hand, have argued that it was flawed in many ways. We cannot then accept Birch's standard: it permits scientists to exercise influence through a kind of self-validation. We seek governance standards that permit evaluation by *others*. A self-applied test for the permissibility of speaking cannot do the work of grounding political authority for those utterances. Birch does not claim it can, so this is not necessarily a criticism of his account, but it means that the account does not answer the problem that the Pandemic posed for scientific advice in emergencies, at least as we see it. The problem was not whether scientists could speak, but whether they should be followed.

9.4 Indirect Political Authority

We now turn to our proposal for a framework that specifies when scientific advice may acquire a kind of political authority, not through institutional power or public deliberation, but through the epistemic structure of the situation: an account on which scientists may have political authority because of what they know.

We say that *indirect political authority (IPA)* exists when:

(a) A person, institution, or other actor has exclusive epistemic access to a domain, and knows a certain fact in this domain [Exclusive Epistemic Access];
(b) There is a background of widely accepted knowledge and values to which no reasonable subject could object [Shared Background]; and
(c) Combining this fact with this background greatly restricts the range of acceptable policy options, potentially to a single one [Keystone Fact].

"Fact" should be read as including the conjunction of multiple facts.

Illegitimate IPA then arises when a person or institution is granted the rights and powers appropriate to legitimate IPA, but at least one of (a)–(c) is not satisfied.

These conditions capture the situation of Hooper and give structure to the intuition that he was justified in his actions. First, consider whether Exclusive Epistemic Access is satisfied. Hooper is the only marine biologist on Amity Island. On the island, he is the only person with epistemic access to the domain of marine biology. The facts in question concern bite patterns, shark behaviour, and other things falling squarely within Hooper's domain of expertise. We are led to believe that Hooper does indeed know these facts. Nobody else does. So, Hooper has exclusive epistemic access.

Clearly, a contextualist understanding is required here: the relevant context in *Jaws* is Amity Island. David Lewis writes:

> If I say that every glass is empty, so it's time for another round, doubtless I and my audience are ignoring most of all the glasses there are in the whole wide world throughout all of time.
>
> *(Lewis 1996, 553)*

Likewise, when we say that only Hooper has exclusive epistemic access to marine biology, we are not denying the existence of other marine biologists. It is true that more could be called in to give second opinions. It is true, too, that Hooper could do more research to render his evidence stronger. Neither fact undermines the claim that, for present purposes, Hooper knows, and only Hooper knows, that a dangerous shark is in the local waters. Given the time pressure, the nature of the threat, the strength of the evidence, the level of expertise and experience of Hooper, we can say that he knows.

There is no mileage in seeking to theorize these factors in advance. How could one specify, in abstract, how sure Hooper must be, how much disvalue we attach to a shark attack, and so forth? Theorizing these things is difficult, perhaps impossible, as is often the case with contextual factors. People involved in a given context must decide whether, in all the circumstances of the case, and on whatever relevant standards may exist, whether formal and informal, through whatever process of discussion, reflection, negotiation, might be relevant, whether the standards are satisfied. Below, where we discuss governance, we offer some guidance as to the structure of this discussion. But it would be unhelpful to attempt a kind of theoretical legislation in advance. When statutes are written, if they are good statutes, they leave certain things to be thrashed out by the parties involved, recognizing that it is impossible to foresee all possibilities. Actual outcomes will always depend on individual and collective competence and honesty.

The Shared Background condition is also satisfied in Hooper's case. The implication of the film is that there is a shared background covering both facts and values, and, if you like, any combination of the two, concerning the desirability of getting attacked by a shark, and how to balance this risk against the benefit of a nice swim. This background could include legal duties – for instance, a duty on the authorities to prevent the public from accessing a beach if it is dangerous to do so. Or it could simply include moral duties of a very strong kind, universally shared by reasonable subjects and not such that a reasonable subject could object. Reasonable people often do agree that authorities are morally obliged to exercise their legitimate powers so as to protect people from imminent serious danger, for example. So it seems there is a shared fact–value background of the right kind.

The Keystone Fact condition is supposed to capture the idea that these circumstances are satisfied. In the *Jaws* case, we are invited to think that they are. The presence of a dangerous shark morally obliges the Mayor to take action, and his refusal to do so is wrong.

Why does this confer on Hooper the right to *tell* the Mayor he ought to close the beach? Because in doing so, Hooper is not claiming any special knowledge outside his domain. He is merely connecting the special knowledge he possesses with a fact–value framework that *everyone* possesses (or every reasonable person). This includes the Mayor. He is not making a recommendation about what he thinks should be done; he is saying what *any* reasonable person would agree should be done, if that person knew what he knows.

You may yet ask: if the Mayor would agree if he only knew the facts, why should Hooper make a recommendation or lobby the Mayor rather than merely lay out those facts, and allow the Mayor to agree? Because understanding those facts is beyond our Mayor. Hooper cannot convince the Mayor of the facts. This is because the Mayor is not an expert (perhaps tinged with bias) and does not understand the domain. Hooper does not have the time to educate the Mayor in marine biology. (And anyway, there is no guarantee the Mayor, a politician, would be sufficiently cognitively capable.)

In this situation, we suggest it is at least morally acceptable and probably morally required for Hooper to try to persuade the Mayor to close the beach. This requires a suppressed premise to the effect that it is morally acceptable or required to tell an authority when they are morally (or legally) required to do something. It seems that society generally operates as if that were true; thus, it is, in turn, generally morally accepted and part of the background condition. Hooper does not do wrong; at least, reasonable subjects would agree so, and no subject could reasonably object.

The key point is this: Hooper is not exploiting his position to share his own opinions. He is saying what any reasonable person would say if that person knew what he, Hooper, knows. Thus, there is no problem of undue influence. Provided Hooper is indeed referring to a shared background and is not mistaken about it, he does nothing wrong by pointing out what the shared values dictate, given the facts to which he, Hooper, has exclusive epistemic access. To put it another way: Hooper was not dictating policy, merely telling the Mayor that anyone would agree needed to be done, if they knew what he (Hooper) knew.

Thus, we get the result we desired. Hooper has no direct political authority in virtue of his being a scientist. Rather, it is Hooper's knowledge that grants him the right to seek to influence policy. The only way in which his values play a role is to the extent that they are shared with everyone else (who is reasonable). He has political authority not in virtue of his being a scientist, but in virtue of what he knows. This is the result that the many audiences of *Jaws* have intuitively accepted.

9.5 Pandemic IPA

Scientists who sought to influence policy during the Pandemic may, at times, have seen themselves as in the same boat as Hooper. We have shown in earlier chapters how predictions about the spread of Covid-19 played a central role in policy decisions during the Pandemic, particularly in relation to lockdown. We have also shown how some scientific advisors sometimes went beyond merely publishing these predictions and took various actions to influence policy. They even went so far as to act against those whom they believed to be mistaken – those "fringe epidemiologists" who, they felt, lacked epistemic access, or whose judgement may have been clouded by bias, as the Mayor's judgement was clouded by the potential for lost profits. All this is hardly surprising: it would be psychologically impossible to remain neutral between options if some were catastrophic. IPA offers a framework in which this *might* be justified. If the scientists in question had possessed this framework, it could have saved them a good deal of floundering in the UK Covid-19 Inquiry, where they were repeatedly challenged about the tension between their stated position (that science just offers facts), and evidence from emails that showed the expression of opinions about policy options.

We will now argue, however, that the conditions for IPA were not satisfied by those modellers who sought to exercise influence during the Pandemic. Here we focus, again, on the influence of ICL and the UK's SAGE. There were others, but we have already given our reasons for thinking that these had a global influence, and they illustrate the argument well.

9.5.1 Exclusive Epistemic Access

So first, did infectious disease modellers have Exclusive Epistemic Access to a relevant fact? No, they did not. There are two reasons for this, and the first is specific to the peculiar dynamics within that field at the time. It is the methodological rigidity that dominated the field at that time. As discussed at length in Chapters 2 and 3, this led to some modelling methods (mechanistic approaches) being deployed rather than others (descriptive approaches) that would have been more appropriate, or at least served as a sense check, prospectively and retrospectively. We have already given our arguments that those teams persisting with a mechanistic approach without evaluating its performance made some extremely large errors, at the expense of their credibility. These errors imply that they did not know the facts they claimed to know, and therefore, Exclusive Epistemic Access was not satisfied.

The second reason that these modellers failed Exclusive Epistemic Access is that predicting the outcome of a public health intervention calls for input from multiple disciplines. This point does not depend on complexities about the kind of disease modelling that was being used. It is public health 101 that one size does not fit all (Broadbent and Smart 2020; Smart and Broadbent 2020). Facts from anthropology, economics, sociology, and many other fields are going to be relevant to predicting what will happen when you make an intervention (see Figure 9.1). As Nancy Cartwright famously points out, just because a policy works somewhere does not mean it will work for you (Cartwright 2010). It is by now obvious that many relevant factors were ignored in many Pandemic policy decisions in many places.

This is now widely acknowledged, including by people centrally involved in those decisions. For example, Graham Medley was chair of the UK Scientific Advisory Group for Emergencies modelling sub-committee (SAGE SPI-M), which advised the UK government during the Pandemic. During testimony to the UK Covid-19 Enquiry, he confirmed that he never received an answer to the question, despite asking it more than once, of who was supposed to be modelling or exploring the impact of measures on the economy, education, mental health, and societal well-being (UK Covid-19 Inquiry 2023a, 112–13). It follows that, as public health advice, SAGE's advice was inadequately informed.

The phenomenon of ignoring relevant information was particularly obvious concerning the effect of lockdowns on the global poor. The effect of a stay-at-home order in a leafy suburb will differ from the effect in a slum, as we show in Chapter 10. There, we make an example of ICL12, which modelled the

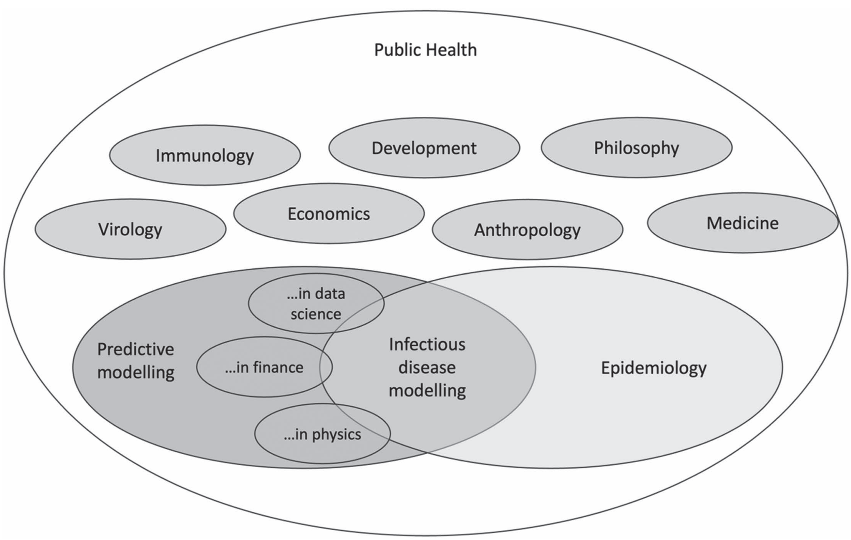


FIGURE 9.1 Disciplines relevant to the public (partial representation, no scale)

effect of NPIs in terms of protecting healthcare systems that in many places do not exist, are not sufficiently capacitated, or are not accessible. Over half of deaths globally (outside pandemic times) occur outside hospitals, and in low-income countries, that is 80% (Adair 2021). ICL12 also understood lockdowns to reduce social contact by 75% across all age groups, which, for the billion people living in a slum, it does not. These assumptions were made out of necessity; that is, because the modellers did not know the facts. There were relevant domains of expertise besides theirs; they lacked exclusivity.

What *could* have been modelled with mechanistic models was the effect of reductions in social contact on disease transmission. What was, in fact, modelled was the effect of various policies. But policy formulation and evaluation are specialized professional areas (Whitty 2015). When they sought to model the effects of non-pharmaceutical interventions, infectious disease modellers were going well outside any domain to which they could possibly claim exclusive epistemic access. This does not mean they should not have done any modelling: it means they should have been more humble about their projections, and, rather than attempting to shape policy on their basis, should have sought and accepted that the inputs of many other domains of knowledge would also be relevant – not just for deciding what to do, but even for predicting what would happen. Sometimes, some of them did not communicate this limitation on their knowledge.

9.5.2 Shared Background

To exercise legitimate IPA, there must be a background of shared knowledge and values that no reasonable person would reject. Was there such a background? Again, we contend that there was not. This is obvious because everyone knows that many people did indeed object to lockdowns on the basis of a lack of shared values. There were debates about liberty rights and about the weighting of the interests of different groups (age, income, etc.). The common response is to argue that urgency dictated swift action. However, this is circular: urgency dictates swift action only to the extent that this action is one that reasonable subjects would agree on. This does not mean what they would agree *on in ideal circumstances*. It means: would agree on, *given the urgency*, if they knew what the scientists knew. It is obvious that this was not the case.

Consider another situation. When disease breaks out among domestic animals, it is common to cull many healthy individuals in order to forestall wider spread. Suppose that, instead of quarantine, the authorities in Wuhan had instituted a cull of the entire population of the city, thinking it would prevent the spread. This is obviously not a good public health measure. But *why* not? Culls and lockdowns share the feature of sacrificing a smaller number of lives if doing so saves a larger number, even though killing is a necessary part of the cull, but an unfortunate by-product of the lockdown.

We would suggest that it is a part of the shared background that culls would be morally unacceptable as a response to Covid-19. Critical as we might be about many lockdowns in response to Covid-19, we certainly do not regard them as morally equivalent to culls. However, culls and lockdowns have these similarities: both involve prioritizing one aspect of well-being over others, and both involve prioritizing the interests of some people over those of others. It is natural to say that culls are not justified because reducing Covid-19 mortality does not justify killing people (nor even letting them die in a quarantine zone). However, lockdowns kill people too; the difference is therefore one of degree, not kind. If one merely points at the number of lives involved, then the most effective measure would be the most justified, be it lockdown or a nuclear strike.

The framework for justifying lockdown, therefore, needs to have more in it than just saving lives. It ought to make it not just possible but *easy* to explain how culls differ from lockdowns. But in most places, no such framework was deployed, and international organizations such as the *WHO* gave no indication of how to balance either different dimensions of well-being or the interests of different groups. There was no systematic approach to assessing the acceptability of different trade-offs (Broadbent 2020).

This gap is one important reason for the painful politics of the Pandemic. People highly valuing certain principles, notably freedom, were forthright in pointing out that Pandemic response measures did not seem to respect those values. They were right. The background of principles assumed by Pandemic responses was often not sufficiently widely shared so as to qualify the informing science for IPA. This is not to say that the decisions were wrong. It is simply to say that they were open to debate. The debate could have come down on the same side as the decision. But the fact that there was a possible debate about, for example, how to weigh the interests of schoolchildren in education against the interests of the elderly in avoiding infection (or, if you prefer, how to weigh up the public interest in each) implies that reasonable persons could reject some of the possible conclusions of this debate. The Background condition for IPA was not satisfied.

It is unlikely that scientists themselves will be reliable judges of whether the Background condition is satisfied, particularly in relation to values. Scientists are often atypical members of their societies, not only in terms of epistemic expertise, but also in their cultural, political, or moral assumptions. As a result, they may be out of touch (sometimes amusingly so) with the values and priorities of the broader public. For this reason, we do not envision the IPA framework as a self-governance model in which scientists both claim and assess their own authority. Rather, we take seriously the possibility that determining whether the Background condition is met may require oversight by others, especially those better positioned to assess prevailing public values. The IPA framework identifies when scientific authority may be justified, not who gets to declare that it is.

9.5.3 Keystone Fact

During the Pandemic, there was no factual keystone to place in an arch of existing knowledge and principles. There were too many other stones missing. Given the foregoing discussions of exclusivity and background, this is an unavoidable conclusion.

The facts that were treated as keystones concerned the spread of Covid-19 and consequent mortality. However, even if the Orthodoxy's access to them had been adequate, they were not keystones. Even if a disaster of the scale foretold by the most pessimistic projections was approaching, this fact alone was not sufficient to complete a policy "arch." There were too many unknowns about the consequences of the various policy options, both concerning the course of the disease and its consequences in other respects. This is a consequence, as stated, of the arguments above. The predictions emanating from the Orthodoxy certainly created an imperative to take action, and no doubt exerted a significant constraint on policy choices. But before long, locking down had become the only acceptable policy option in many places. The facts, according to the Orthodoxy, did not require this level of restriction. Indeed, this is even clearer in some of the early modelling, where lockdown itself was never modelled, only suppression *short of* lockdown (Ferguson et al. 2020). The fact that such reports were taken to support lockdown is therefore a curious one, indicating the positive feedback tendency of an orthodoxy.

9.6 IPA in Governance of Science

The idea that scientific research should be "governed" may sound alarming because it suggests restriction and interference. This could be detrimental to the pursuit of useful knowledge, and it could provide a route for malicious or misguided influence. However, some scientists and scientific institutions are advocating for science to become more politically active. *Nature* indicated in October 2020 that it would be "publishing more politics news, comment and primary research in the coming weeks and months" (Nature Editorial 2020), and in *Science*, Agustín Fuentes has argued in a series of pieces that science should play a more active role in shaping policy and society (Fuentes 2024b, 2024a). If so, then science needs governance structures appropriate to its political role, to prevent power being wrongly claimed and used, to preserve integrity, prevent bias and infiltration, ensure accuracy, and so forth. Good governance structures within science will protect what is valuable about science, but they will also prevent it from having more influence than it should. Nobody thinks science or scientists should rule the world. The point, rather, is for science to have *appropriate* authority in *relevant domains*.

Developing structures that are adequate for the role of science in contemporary politics will take time. We offer IPA as one contribution to this larger development. The simplest way to implement it would be as a protocol for scientific advisory groups for emergencies, which exist or are convened in most places. The UK's SAGE, for example, could have a protocol that its advice to the government be accompanied by a self-assessment as to whether this advice satisfies the three conditions outlined. The ministers receiving the advice could then decide whether this is convincing. Scientists outside SAGE could, too.

For example, in the disastrous report of December 2021, if SAGE had sought to explain why the two models whose results it published and upon which it based itself provided exclusive epistemic access, they would have needed to say something about the irrelevance of other models, or alternatively, that they agreed. This would have been difficult because other models were relevant and were not used. SAGE could alternatively have acknowledged that their information was only partial, in which case, it is unlikely that the report would have created the impression of a strong recommendation to significantly tighten restrictions. Either way, matters would have been better.

The Shared Background condition would have been even more useful, since few scientific advice bodies anywhere sought to produce any evaluation of the value framework within which they were operating. This was generally said to be not part of their job. And yet they *were* operating within a value framework. Whether or not this is a necessary part of science, as some philosophers contend (Douglas 2009), it is a necessary part of formulating scientific advice, as Pamuk argues. This is why nobody modelled the effectiveness of an early cull of the entire population of Wuhan. Value judgements inform the options a scientist presents to a policy-maker, however impartial that scientist seeks to be; and, as we have argued, it is not reasonable to expect scientists to be impartial when they believe the stakes are extremely high. In such situations, explaining why the stakes are high, not just in their eyes, but in the eyes of any reasonable person, would help the rest of us decide whether the scientists are right about what the rest of us think. In the Pandemic, the background values were far more complicated than the Scientific Orthodoxy implied. And, in our view, it should have known this, if only from the fact that no prior pandemic planning had considered lockdowns likely to be widely accepted. But even if their mistake was reasonable, the expression of what they took to be the shared background of facts and values would have supported political and public deliberation. Urgency might not permit deliberation, but urgency only precludes deliberation to the extent that one can suppose a shared background, and if an urgent decision involves a value judgement, then a politician is the right person to make it, not a scientist. This is why the decision to go to war is made by politicians and not soldiers.

9.7 Conclusion

In contemporary society, science will inevitably possess a degree of political authority in some situations, due to its exclusive epistemic access to certain facts. Moreover, although science does have governance structures, these are not and were never intended to be suitable for conferring political authority, at least not to the extent that contemporary societies will sometimes need or want to confer. The governance structures of science are designed to support the pursuit of intellectual goals and are not able to bear the weight of political authority. If this analysis is correct, we need to develop new structures that are more suitable for this purpose.

These structures will concern a much wider situation than the one that exists when IPA is claimed or attributed. They will concern synthesizing inputs from multiple advisors and evaluating the upshot. There are now calls to develop mechanisms for these processes. Our contribution is to provide a test for the attribution of IPA. If IPA is claimed or attributed, that claim or attribution must be tested against our three conditions of exclusive epistemic access, unobjectionable background, and keystone fact. If it fails these, then it may still be relevant. Clearly, ICL9, ICL12, and other similar models were policy relevant. But they were treated as if they had IPA, when they did not. If the test for IPA had been applied in the manner we describe here, this could have been avoided, without compromising the need for urgent action.

10

SCIENTIFIC INJUSTICE IN THE COVID-19 PANDEMIC

10.1 Introduction

This chapter introduces the idea of *scientific injustice*. By this, we mean a specific kind of wrong that occurs when scientific authority shapes policy, and is negligent in doing so, with respect to the distribution of the benefits and burdens of that policy. This chapter focuses on a specific policy, lockdown, and in particular, the period during which lockdowns were globally enacted during the first half of 2020. Lockdowns, we contend, were a direct consequence of the intervention of powerful scientific authorities. They distributed their benefits and burdens very unequally. This was foreseeable and was, in fact, foreseen by many. And, we argue, acceptable alternatives to lockdown existed that distributed burdens and benefits more equally. On these four bases – scientific influence, unequal outcomes, foreseeability, and avoidability – we construct the argument that scientific injustice occurred during the Pandemic.[1]

In Section 10.2, we define the notion of scientific injustice. It is a form of social injustice, where that is understood in terms of unfair and unequal distribution of outcomes, and the nature of the wrong is negligence: failure to take care. Section 10.3 shows that science-guided policy during the Pandemic, by focusing on two particularly influential interventions by the *WHO* and Imperial College London (ICL) modelling team, respectively, in early 2020, and using these to illustrate the high degree of constraint under which national decision-makers were operating. Section 10.4 shows how the burdens of lockdowns fell much more heavily on the poor, and Section 10.5 shows that the benefits, too, were unequally distributed, with wealthier people and areas tending to benefit more. Section 10.6 shows that this was all foreseeable and avoidable. It was

DOI: 10.4324/9781003732495-11

knowable in advance that locking down was not the best policy in many contexts globally. Indeed, it was known, but those who knew it found themselves unable even to say it, at times, let alone to shape policy. And alternative policies existed, which would have bought most of the benefits of lockdown for a much lower social price. Finally, we consider how scientific injustice relates to the rest of the framework. We consider the possibility that injustice is not a necessary component of orthodoxy properly construed, but rather an additional ethical condition, or perhaps an aggravating factor. We argue, on the contrary, that negligence-type injustice is inevitable when the other four elements of scientific orthodoxy are present, and therefore that the inclusion of this condition in the definition of scientific orthodoxy (and not as a mere aggravating factor) is well-motivated. Ultimately, it offers a framework that makes more sense of the Pandemic.

Here, as throughout the book, we are not condemning the whole of science, but focusing on a particular dynamic (scientific orthodoxy) by which certain scientific authorities guided the widespread adoption of particular policies. Within this scope, lockdowns are particularly important. The global spread of lockdown in early 2020 is perhaps a unique event in the history of science; we cannot think of another like it.

Many scientists see themselves as committed to a fairer world. That makes it all the more striking if science, at a moment of peak influence, brought about a less fair world, compared to an available alternative. It may seem puzzling, even implausible, that science intervened to make things significantly less fair than they would have been without that intervention. The way we characterize scientific injustice in terms of negligence is intended to help make sense of this. The implausible claim is true: the interventions of some scientists did result in significantly unfair policies relative to viable alternatives – but this was, essentially, an accident. However, it was a morally significant one, arising from a failure to take due care: it was negligence.

10.2 Defining Scientific Injustice

Scientific injustice occurs when

(a) Science guides policy, which
(b) Distributes benefits or burdens unequally, where this is
(c) Reasonably foreseeable and
(d) Avoidable

Scientific injustice is thus a form of social injustice as it is usually characterized, that is, in terms of inequality of outcomes.

It might be objected that our definition is too broad, and that some unequal distributions of benefits and burdens are justifiable. Some of these objections can be defused by a proper discussion of the concept of equality, something we cannot embark on here. But there remains a live question about whether equal distribution of benefits and burdens is an intrinsic good, another about whether social justice is best defined this way, and a further question about whether social justice is a form of justice at all. We want to avoid all these issues by restricting our discussion to the case at hand. We work from a general premise that science is no more intended to benefit one person than another, and that its benefits ought ideally to accrue equally to all; and, conversely, that any burdens it creates should be equally distributed. It follows from this working premise that policies guided by science should be as equitable as possible, since if they are not, then science is benefiting some people more than others. We will not defend the working premise here, but would point out that it does not amount to a broad commitment to egalitarianism, merely a default starting point. In the absence of any reason to do otherwise, we suggest cutting the cake equally. We are content to put those who disagree to the proof as to which people science should prefer, and willing to wager that it will be a group including them.

Injustice may arise from an accident, but only if that accident arose from a moral failing. The moral failing we allege is negligence: failing, or neglecting, to take due care. The view is that negligence is a part of common morality in many societies, and is reflected in common law tradition by the tort of general negligence, on which our definition is loosely modelled.

The distinguishing feature of negligence is that it does not place onerous requirements on the subjective mental state of the wrongdoer. A person can be negligent without intending to cause harm, or even knowing that they might do so. The requirement is simply that it have been reasonably foreseeable that their actions would cause harm. The sense of reasonable foreseeability invoked here is that used in the common law and is an objective test, as we explain below.

We now take each of (a)–(d) in turn and show how it was satisfied during the Pandemic.

10.3 Science Guided Lockdown

We say that *science guides policy* when, but for the role played by scientific authorities in shaping policy decisions, those policy decisions would have been significantly different.

The policy in focus in this chapter is lockdown, which we define as follows.

> A lockdown is a legal requirement for all persons to stay at home unless their reason for being outside the home falls within a closed list of exceptions.

We discussed the correct definition of lockdown in Chapter 5, and showed in Chapter 4 that science was instrumental in constructing this policy framework. For the purposes of the present chapter, the exact definition of lockdown does not matter, because most policies implemented under that name are subject to the same line of criticism that we take here.

How did science guide lockdown? As we explained in the *introduction*, a large portion of the book is devoted to answering this question. The mechanism involved methodological rigidity, scientific dogma, suppression of dissent, and indirect political authority. In this chapter, we aim to show merely that it *did* guide the lockdown.

In showing that science-guided policy with regard to lockdown, we inevitably also say something about how it did so, and whose science did the guiding. We contend that a relatively small group of scientific authorities guided lockdown globally. This results in an important further consequence: disregard for local context. Our treatment in this section will focus on two specific scientific interventions that particularly clearly illustrate active efforts to shape national policies around the world, and not merely local national policies. At the same time, we will show that the early global spread of lockdowns was not driven by local factors, including local science, and how this resulted in a mismatch between policies and local context.

We choose these global-facing interventions because we take the international dimension of the scientific injustice of lockdown to be particularly important. We see the rapid global spread of lockdown as the widespread dissemination of a policy designed by and for wealthy people in wealthy places to completely different contexts as at the heart of what we see as injustice. The continuation of imperial power structures in science is particularly resilient in epidemiology and public health (Campos-Rudinsky et al. 2024), as we discuss in Chapter 2. The two most powerful institutions in the Pandemic were ICL and the London School of Hygiene and Tropical Medicine, and the historical power structures reflected in their names may have changed but have not disappeared. This is an important part of the explanation for the negligence we allege.

10.3.1 WHO China Joint Mission Report

On 19 February 2020, the *WHO* began an official mission to China, issuing a Joint Mission Report five days later. This document contains substantial scientific input. Much of it is virological, but the latter part concerns policy. China is strongly commended for its response to Covid-19, including its lockdown measures. This commendation is significant not only for its praise but also for its form: it implicitly recommends emulation. This simultaneously illustrates our core point – that science-guided lockdown policy globally – and our core concern with this guidance: that it made a global recommendation without due regard for local context.

"Recommendation by praise" is an interesting diplomatic tool. It has the appeal of being non-directive (so compliance is not required), of showing that others have already acted (so action seems feasible), and of quietly shaming the laggard (so inaction feels humiliating). The report makes no explicit recommendation to implement lockdowns, but it strongly approves of China's response. For example:

> China's uncompromising and rigorous use of non-pharmaceutical measures to contain transmission of the COVID-19 virus in multiple settings provides vital lessons for the global response.
>
> *(WHO 2020a, 19)*

These lessons include the following:

> extremely proactive surveillance to immediately detect cases, very rapid diagnosis and immediate case isolation, rigorous tracking and quarantine of close contacts, and an exceptionally high degree of population understanding and acceptance of these measures.
>
> *(WHO 2020a, 19)*

But not all countries can do these things. The report also emphasizes the role of AI and the diversion of significant resources. And not all countries have these things.[2] The *WHO* China Joint Mission Report emphatically provided no alternative to China's approach. For an organization with global responsibilities, this was irresponsible. In this situation, nations had nowhere to turn for independent guidance, and it is unsurprising that the strongest predictor of national policy was the policy in nearby nations (Sebhatu et al. 2020).

China is one of the largest economies in the world. It has an extremely strong state which can take major decisions centrally, and fast. It is not dependent on foreign aid. It is enormous, geographically and in population, and encompasses a wide variety of living contexts, and most of these were not placed under a lockdown. The approach was to quarantine certain areas and populations. These exceeded the size of many nations, but left many other parts of the country largely untouched, and afforded opportunities for internal support that would simply not exist for a country of, say, five million. (Wuhan, the city where Covid-19 emerged, had a population of about twelve million in 2020, excluding the larger metropolitan area.) In short, China is an unusual country, and unusually well-placed to pursue a policy of extremely rigorous social distancing. Smaller and especially poorer countries would face a very different set of considerations.

The Joint Mission Statement incautiously accepts the Chinese claim that life in China was returning to normal. This compounds the failure to consider international differences of context. A key ingredient of any national deliberation on response measures would be the duration for which they would be needed, and whether they would need to be reintroduced. The time a policy would need to be in place would be a major determinant of its cost, in both economic and human terms. A message of precaution would have been appropriate, both for influencing policy elsewhere and for preparing people for the possibility of a long haul. On 24 February 2020, it was far too early to assert that life was returning to normal. The Chinese state may have had an interest in conveying the impression of success. The *WHO*, however, ought to have perceived this and to have balanced it against the interests of all other member nations in accurately evaluating the probable costs of the approach taken in China.

The report contains strong evidence of Chinese state influence. The Major Findings section includes overtly political elements that lack any relevance to public health. For example, instead of focusing on the way that response structures were organized, it emphasizes the role of national leadership in directing the response:

> General Secretary Xi Jinping personally directed and deployed the prevention and control work and requested that the prevention and control of the COVID-19 outbreak be the top priority of government at all levels. Prime Minister Li Keqiang headed the Central Leading Group for Epidemic Response and went to Wuhan to inspect and coordinate the prevention and control work of relevant departments and provinces (autonomous regions and municipalities) across the country. Vice Premier Sun Chunlan, who has been working on the frontlines in Wuhan, has led and coordinated the frontline prevention and control of the outbreak.
>
> *(WHO 2020a)*

The Chinese state had powerful reasons to want to present itself well in the eyes of the world, due to criticism of its handling of the initial outbreak. The initial handling of the Pandemic by China and the role of the *WHO* in downplaying the risk of widespread human-to-human transmission are important but not directly relevant.

Unfortunately, the result of this indirect approach to outward-looking recommendations and an overly positive assessment of China's response was that the Joint Mission Statement generally recommended a policy that was particularly well-suited to China, without consideration of the many ways that China is unusual or even unique.

The *WHO* is a hybrid body, not purely a scientific one, and is always treading a delicate political line. However, its advice in such matters is widely supposed to have scientific backing. The scientific position it presented, however, had no regard for consequences in places and for people without the means to make their own assessments, and for whom the proposed measures would foreseeably cause much greater hardship. The report is good evidence that science-guided policy globally, since it is plausible that this report was globally influential; but unfortunately, it is also evidence that the policy was locally inappropriate in many places.

10.3.2 Imperial College London Report 12

The Joint Mission Report recommended emulating China, but was not specific. A more direct piece of scientific guidance that shaped global responses to lockdown came from ICL Covid-19 Response Team. The most famous of its 51 reports was Report 9, discussed in detail in several chapters of this book. Report 9 predicted the outcomes of various response strategies in the United Kingdom and the United States. Report 12 (ICL12), however, produced global projections. This again illustrates our core claim – that science shaped lockdown policy – and our core concern, which was that there was inadequate attention to local context. ICL12 did not take local circumstances into account, as may be imagined in a global exercise of this magnitude. This required it to make some strong assumptions. Unfortunately, these assumptions were too strong and were not always recognized.

Two stand out in particular. First, in common with Report 9, ICL12 focuses on hospital bed capacity. The rationale of the modelling of this era was to protect healthcare systems from being "overwhelmed." However, in low-income settings, protecting hospital capacity is a much less obvious priority because access and level of care are commonly low. Over half of all deaths globally occur outside hospitals, and in low-income countries, the figure is 80% (Adair 2021). Protecting hospital beds is less relevant in a context where people with terminal health problems do not usually access a hospital bed.

ICL12 made "the strong assumption that similar levels of medical care to that provided in China are available elsewhere" (Walker et al. 2020, 15). This significantly misunderstands the global healthcare situation. There are countries with no ventilators at all, and hospitals where paracetamol is prescribed for malaria. Accessing such hospitals with a case of serious Covid-19 is unlikely to greatly improve prospects of survival.

Concerning the question of wider applicability, the report says:

> These assumptions may mean that our results may overestimate mortality in some HICs and under-estimate it in some lower income countries.
>
> *(Walker et al. 2020, 15)*

This phrase makes it sound as if the assumptions are conservative and that making them more precise would result in an even stronger case for action in low-income settings. In fact, it means the opposite. The mortality in question is mortality following various interventions. What this sentence means, then, could equally be expressed by saying that in low-income settings, these interventions are likely to have a smaller beneficial effect.

The second stand-out assumption in Report 12 is not identified by the authors, which is that strategies would have the same effect on social contact everywhere. Strategies were modelled by social contact rates, without consideration of whether different strategies would result in different contact rates in different countries. Thus:

> We also explored the impact of more rigorous social distancing approaches aimed at immediate suppression of transmission. We looked at 6 suppression scenarios in which the timing of policy implementation varied according to when the weekly death rate per 100,000 population exceeds a certain threshold (here, either 0.1, 0.2, 0.4, 0.8, 1.6 or 3.2 deaths per week per 100,000 population) the effects of widespread transmission suppression were modelled as a uniform reduction in contact rates by 75%, applied across all age-groups.
>
> *(Walker et al. 2020, 16)*

Thus, "rigorous social distancing" is modelled as a reduction in social contact by 75%. Below, we will argue that this is unachievable in slums, even under a strict lockdown. About one in eight people lives in a slum. This must cast doubt on the validity of this global modelling exercise. Similarly, there is no modelling of the gap between a regulation's content and its effect on social contact. A recommendation in Stockholm might reduce social contact more than a stringent lockdown in Soweto, for a number of reasons, discussed later in the chapter.

This gap between regulatory intervention and effect on social contact runs through all the epidemiological modelling we discuss in this book. Modelling the effect of an intervention requires making assumptions about how the regulatory or other action taken by the authorities affects the parameters of your model. Will people obey a lockdown? Will they act on a recommendation? To what extent, and how much of a reduction in social contact will this yield? We have argued in Chapter 9 that epidemiological modellers went well beyond their field of expertise in trying to answer these questions. Here, the point is simply that their answers were assumptions, and that these assumptions did not reflect the circumstances in which much of the global population lives. Scientists were guiding policy globally, but the policy reflected the scientists' local context.

10.3.3 Local Decisions Were Constrained

One might object that local governments took decisions during the Pandemic, and that the responsibility for these decisions thus rests on them, and not on scientists producing reports on the other side of the world. This is less convincing as regards the *WHO*, whose job it is to inform member nations about health threats and strategies. Nonetheless, the *WHO* has no control over its member nations' policies. One might even argue that we are denigrating nations by denying their agency and portraying them as unduly weak.

This line of objection overestimates the degree of autonomy that national governments enjoy. Evidence that NPI policies were strongly influenced by the policies adopted in other countries was available in 2020 already. One study of *OECD* countries found that democracies are more prone to this tendency, while slower to impose restrictive measures (Sebhatu et al. 2020). A more extensive and methodologically sophisticated study in 2022 finds that "peer effects, particularly mimicry of geographic neighbours, political peers, and language agnates, drive policy diffusion and shape countries' policy choices" (Mistur, Givens, and Matisoff 2023). In other words, countries tend to do what other countries do.

From a strategic perspective, this is unfortunate, of course, because:

> Following the lead of others rather than making decisions based on the specific situation of the country may have led to countries locking down either too early or too late. Conversely, if countries follow each other when easing restrictive policies or reinitiate such policies, there may be a situation where countries adopt epidemiologically suboptimal policies.
>
> *(Sebhatu et al. 2020, 21201)*

There are obvious political dependencies between countries. Many low-income countries depend significantly on international aid. In 2022, Oxfam reported that international aid funds the entire social protection programmes in seven countries in Africa (Oxfam 2022). In some countries, the level of aid has at times amounted to over 100% of total government spending, though the level of reliance has been falling in recent years (Bwire 2023). The *WHO* China Joint Mission Report would have had an effect on governments receiving Chinese investment, not only because of the status of the *WHO*, but also because of the importance of Chinese investment, notably in Africa, where most low-income countries are located. ICL12 does not carry its own direct political power, but it will have influenced scientists, because the power dynamics in international science reflect political power dynamics.

Scientific power, like political power, is strongly concentrated in global north institutions, and this means that scientists are also constrained in what they say about advice coming from those institutions. In the first place, replicating or competing with the work of those institutions is very difficult for small teams in low-resource settings. The type of modelling used to generate the influential international regulations requires specialized expertise of several kinds, implying large teams, and requires computing equipment unavailable in most countries, low-income or otherwise. Even if this difficulty were overcome, accessing international funding for research typically depends on collaboration with a large Global North institution, and this provides a disincentive to question strong statements issued by potential future collaborators.

There are further power relations built into the epidemiological modelling and public health community that put global north institutions in charge, as we discuss in Chapter 3. The *WHO* has only one advisory body on infectious disease modelling, namely, that at ICL, and its terms of reference include not just doing research for the *WHO* but also developing and teaching methods. This implies a considerable degree of influence over the attribution of skill and expertise. In Chapter 3, we see that this indeed played a role during the Pandemic, and that scientists who used different methods and reached different conclusions were publicly disapproved.

In no way do we mean to denigrate scientists based in low-income settings. But the prospects for them to conduct alternative modelling and make alternative recommendations, contradicting those coming from the *WHO* and ICL, were minimal.

10.4 Unequal Distribution of Burdens

We say that a policy *distributes benefits or burdens unequally* when some persons enjoy larger benefits or bear heavier burdens flowing from the policy than others. We will begin with the burdens of lockdown.

There are an enormous number of studies detailing the impact of Covid-19 response measures. There is no consolidated assessment, so far as we are aware, and we are not providing one here. A truly comprehensive assessment would be a major undertaking in its own right and may be an unattainable ideal. Also, the headline ought to be obvious: of course, lockdowns are harder in slums, on people without secure jobs, on people with no savings or assets, and so forth, and evidence to this effect emerged quickly (Nyashanu, Simbanegavi, and Gibson 2020). Nonetheless, it is useful to itemize some of the dimensions of inequality and to identify some illustrative factual information. That is what we do here.

10.4.1 Health Burdens

Lockdowns caused health burdens directly through the impact on access to other health services.

For example, maternal and perinatal outcomes were significantly impacted with "considerable disparity between high-resource and low-resource settings" (Chmielewska et al. 2021). Another direct health effect of lockdown was due to the impact on immunization programmes. *UNICEF* flagged the risk to children in May 2020, estimating that 80 million children under the age of 1 were affected (UNICEF 2020). In 2022, a measles outbreak was reported in Africa (Nchasi et al. 2022). There was no measles outbreak in Europe. Measles is a much more serious disease than Covid-19, especially for the young, and much of the population of Africa is young. There was also widespread disruption to treatment programmes for HIV and TB. For example, a study of the impact of the Pandemic on TB indicators in Ethiopia found a significant (negative) impact (Arega et al. 2022). HIV and TB are not prevalent in high-income settings.

The effects of lockdowns on health are also indirect, because lockdowns place other burdens that have a health impact. Thus, the effect of lockdowns on livelihoods affects nutrition; the effect on education affects future health prospects because of the link between educational attainment and health; and so on. These are often intermingled: for example, the above-cited systematic review points out that an increase in maternal death was not only due to lack of access, but also to intimate partner violence, which is a major cause of maternal death and which lockdowns exacerbated (2021, e770).

There are multiple indirect effects of lockdown on health, because the effects on livelihoods, education, and so forth all have health dimensions. The effect on livelihoods is particularly obvious because in low-income settings, economic impacts have major implications for livelihoods, and these have a direct impact on health. Hunger and malnutrition rose sharply from 8.4% to 9.9% in 2020 (UN 2021). The number fell again, but Covid-19 mortality was much higher in 2021, showing (not surprisingly) that the effect on world hunger was more probably a consequence of response measures rather than the disease itself.

This is a tiny sample of the health burdens of lockdown. It is, however, indicative of the way that the burdens fall more heavily on the poor. Measles, hunger, maternal death, and so on are overwhelmingly more of a risk among the poor in any population, and in poor countries compared to rich countries.

Health is not the only dimension of human flourishing. To argue that lockdowns amounted to a cure worse than the disease, as some did, is both substantively and rhetorically mistaken. For some, the health burdens of lockdown plausibly exceeded those of Covid-19. But for others, including some of the global poor, Covid-19 imposed significant mortality. We should not be misread as arguing that lockdowns imposed a greater health burden than Covid-19: even where true, this is not the basis of our assessment. Lockdown placed many other burdens besides burdens on health, and these inhibit important aspects of human flourishing. The message of this sub-section is simply that the health burdens imposed by lockdown are correlated with poverty, and thus that lockdown distributes health burdens unequally.

10.4.2 Economic Burdens and Livelihoods

The spread of wealth in the world is huge. Gross national income (GNI) per capita is below $1,045 in low-income countries and above $12,969 in high-income countries. Gross domestic product (*GDP*) per capita in Malawi, a low-income country, is $1.07 per day, while in the United Kingdom (a high-income country) it is $118 per day. About 9% of the global population lives in extreme poverty, and over 20% live below the upper-bound poverty line, which is still a standard of living that most readers of this book will probably never have experienced. Poverty is very different from its absence, and it is very common.

One effect of lockdown was to make some people poorer. These people tended to be people who were already poor. This is illustrated by the fact that many people fell below the poverty line during 2020, and then rose above them again in 2021 and 2022, despite Covid-19 being more prevalent and more lethal in 2021. The global poverty rate was falling up to 2019. It then shot up in 2020, and fell again in 2021, with 2019 rates resuming in 2022 (Christensen 2023). This was a result of Covid-19 response measures, rather than Covid-19 itself, which was more prevalent and more lethal in 2021 than in 2020. Poverty greatly impacts the ability to withstand a lockdown. Assets, cash, and food stocks in the home are typically lacking. Poverty is associated with livelihoods that are more likely to be disrupted by lockdowns. It is also associated with overcrowding and uncomfortable living conditions.

Lockdown did not make wealthy people poorer to the same extent, and at the top end of the scale, it made some people richer, in some cases considerably so. Some wealthy governments embarked on substantial fiscal stimulus and prompted a stock market boom. Technology companies enjoyed considerable success during lockdown because working from home pushed people to rely more heavily on various technologies, especially for communication. Vaccines must also be considered since they were the (explicit) strategic endpoint of lockdown. Vaccines are highly lucrative, and it was highly predictable that a Covid-19 vaccine would be exceptionally profitable. Oxfam points out that:

> Covid vaccines created 9 new billionaires with combined wealth greater than the cost of vaccinating world's poorest countries.
>
> *(Oxfam 2024)*

The economic cost of lockdown is thus higher in low-resource settings relative to both national and household income, and lower or even negative (i.e. profitable) in high-income settings.

The cause of the growth of poverty was the effect of the Pandemic on livelihoods. Again, this effect was primarily mediated by response measures rather than a direct effect of disease burden, and especially lockdowns. The scale of the informal economy in low-income countries is enormous. In Africa, 80% of the workforce is in the informal sector. This implies no pay for no work. Casual labourers, market or street traders, and agricultural workers will receive money weekly or daily, and generally have no employee benefits or income protection. They are also typically poor.

Workers in developed economies are more likely to have jobs that they can do from home, and more likely to receive support from either their employer or the state if they cannot. Within developed economies, workers are more likely to be able to work from home if they are in better-paid jobs – the kind you can do with a laptop. In wealthy countries, many people found working from home more pleasant, whether or not more productive, than working in the office, prompting a resistance to return to the office, which continues to be a topic of discussion. Many of these people also re-evaluated priorities and chose to de-emphasize work altogether, at least if the zeitgeist is to be believed. This "many," however, is strictly confined to high-income contexts. "Quiet quitting" is a luxury, something that indicates that a level of well-being has been attained that does not require more income and may even be improved by less. This is a remarkable social achievement, but it is not widespread. The vast majority of the global population needs to work to live, and will be able to flourish better in other domains of life if they earn more money. In countries with very high unemployment rates (e.g. South Africa, where the unofficial but plausible unemployment rate is probably over 50%), work is desperately sought, often in despair. Quiet quitting is for the global elite and demonstrates the extent to which lockdown placed its burdens unequally along the dimension of poverty.

In short, the economic burdens of lockdown were generally inversely proportionate to wealth, both individual and national. Sometimes, a tension is posited between economic and health interests. That tension exists only for the very rich. For most people, economic downturn implies harm in all other dimensions of welfare, health included. The poorer you are, the larger this impact.

10.4.3 Education, Children, and Women

Lockdown imposed significantly higher impacts on women and children than on the rest of the global population. The burdens intersect, and so it is helpful to discuss them together.

Lockdown seriously disrupted child welfare in many ways. Education is one of these, because lockdown came with school closures. In many places, children spent significant time out of school, and there was a significant loss of learning, correlated with low income on an individual level (The World Bank, *UNESCO*, and *UNICEF* 2021). Education is a major determinant of the future welfare of a child in health and economic terms (UNESCO 2013). It is the major determinant of employment prospects (OECD 2022). It also plays a role in national development (World Bank 2018). The relationship between wealth and health is well-documented (WHO 2008), particularly in low-income contexts. Low income causes poor nutrition, poor living conditions, and, where there is work, this work is more likely to be hazardous to health, either posing immediate danger of accident or longer-term dangers due to exposure to unhealthy substances and work conditions, or simply by being extremely arduous.

The effect of disrupted education is typically worse for girls, who (like women) were more likely to assume domestic responsibilities during lockdown, and less likely to return to school after closures (UNESCO 2020). Education level is a strong predictor of the number of children a woman will have, being inversely correlated, and the larger the number of children a woman has, the higher the poverty level, since there is less to go around (Götmark and Andersson 2020). (Africa had the second strongest continental economic growth in 2023, but high fertility rates mean per capita growth has been consistently very low (African Development Bank 2024).)

Not all the negative effects of lockdown on children are mediated by school closures or education disruption. Lockdown is very bad for children, even when enacted without school closure (which was, in any case, exceedingly rare). Being able to play football in the street with your friends is very different from having to stay home from psychological, health, developmental, and domestic violence perspectives. For women, who make up the vast majority of primary caregivers, the difference was also significant because lockdown implied greater childcare duties, as these could not so easily be shared with relatives or friends. The ability of both men and women to work benefits children too, since it majorly affects child welfare. Malnutrition is particularly detrimental to children and is a factor in nearly half of deaths among children (WHO 2024d). Thus, lockdown is bad for children, especially girls, aside from its effect on education.

Turning to women, domestic responsibilities in normal times fall overwhelmingly on women. Lockdown intensified some of these, such as child care or meal preparation. The burden of this increase likewise fell overwhelmingly on women (UNSDG 2020). Women are also more likely to work in the informal sector than men, to have lower-paid jobs than men with fewer benefits, to have jobs deemed non-essential (such as secretarial or receptionist roles), and to work in public-facing roles in certain sectors such as retail and hospitality (ILO 2023). Women were thus more at risk of losing their income during lockdown (UNSDG 2020).

Domestic violence is perhaps the most neglected aspect of the disproportionate burden the lockdown placed on women and also on children. Domestic violence includes violence within the home, which is overwhelmingly against women and children. This includes: homicide by intimate partners, overwhelmingly against women; maternal homicide; child homicide; and sexual abuse of both women and children, especially girls. All increased under lockdown (UNSDG 2020; UN Women 2020), though quantifying the increase with any degree of accuracy is impossible.

Although often acknowledged, it is our view that the burden of domestic violence imposed by lockdown was and remains consistently underestimated. Remaining in a space with someone who might hurt, rape, or kill you is not tenable, and Covid-19 did nothing to change this fact. The lack of consideration must indicate a lack of imagination. It seems unlikely that many policy-makers set aside even a short period of time, say three minutes, to visualize a situation in which they faced serious violence if they stayed at home and law enforcement if they left it, which in low-income settings often also implied violence, which could also have been of a sexual nature. Imagine being terrified, perhaps even for your life, but being prohibited by law from running away. The "shadow pandemic" was strenuously flagged by interest groups, but in the wider discourse, this was taken to show that arrangements should have been made to protect women and children from domestic violence. This is simply hand-wringing because it implies a degree of state capacity that is entirely lacking in many countries, and most low-income countries, and a political will that was lacking in most better-off countries, too.

The logical question, then, was: should lockdown be implemented where domestic violence could not be effectively mitigated? A more dramatic, but still logically implied, question would be: should there be a legal prohibition against running away from violence, overwhelmingly impacting women and children? This question was not discussed in the mainstream. The burden of domestic violence was not weighed against the burden of Covid-19 mortality. Yet weighing them up is a non-trivial exercise, because it cannot be settled by a simple calculus of ultimate mortality: rights to personal safety come into play. A lockdown sacrifices a large number of women and children to domestic violence for the sake of reducing Covid-19 mortality. Is that justifiable? It is a live ethical and political question. Yet, it was never a focus of mainstream discussion, even when hands were being wrung. Why not? The answer cannot treat it as a coincidence that the violence was directed against women and children.

10.4.4 Political Stability

Lockdown involves a high degree of state control over individual and group behaviour. In some societies, this raises matters of principle only. In others, it raises a real prospect that the state will misuse its power. This is more common in lower-income countries, though not confined to them. Some countries were strongly resistant to locking down because of a history of state control. South Korea, for example, undertook major responses to Covid-19 but did not lock down, because of a strong, historically inspired commitment to democracy and the rule of law (Campbell 2020). South Korea is a democratic success story, however. Less secure democracies sometimes went backwards. In Uganda, lockdowns were used to control elections in 2021 (Boyd 2023). In Africa as a whole, the number of coups was steady in the years leading up to 2020, but increased significantly subsequently, and lockdowns are certainly among the drivers (Chin and Kirkpatrick 2023).

This risk is another that was not contemplated in wider international discourse, even though it was obvious to people on the ground. Again, this led to important questions being missed. Would lockdown be worth it if it reduced the burden of Covid-19, but led to a dictatorship of indefinite length and appalling brutality? The answer is not obvious, even less so if you are in the context.

10.4.5 Implementation, Enforcement, and Brutality

Lockdowns were implemented with various degrees of competence and force. Regarding competence, a notable example of its absence was the Indian government announcement of a lockdown with minimal notice and no provision for livelihoods, prompting mass exodus from urban areas (Ellis-Petersen and Chaurasia 2020). Another example of incompetence was the South African regulations, famous for irrational prohibitions on a variety of activities such as purchasing clothing and cigarettes, or taking outdoor exercise (Smart and Broadbent 2020).

Well-developed states, with good resources and strong governance structures, are in a better position to develop effective and reasonable regulations. They are also in a better position to mitigate the negative effects of lockdown. Governments in developed countries set aside large amounts of money to pay workers who would otherwise lose their jobs, thereby supporting both the workers and their employers. Less capable governments were less able or sometimes less willing to take mitigating steps, and even where they did, administrations were less able to deliver. To pursue the South Africa example, much of the cash that was intended to support people in need during lockdown never reached them, and the food was also often embezzled, either for sale or consumption. Sometimes administration was simply incompetent, so that there was no control to ensure fair distribution, resulting in long queues, unruly crowds, and some going without. In many cases, there was simply not enough to go round.

Enforcement of lockdowns was sometimes brutal, and this brutality was also not distributed equally. South African law enforcement killed a number of people during lockdown (Harrisberg 2020), and Uganda's lockdown was notoriously brutal, and similar incidents occurred in a number of other African countries (Bujakera and Mersie 2020).

In sum, the burden of incompetent, corrupt, and brutal implementation of lockdown was distributed unequally.

10.4.6 Physical and Psychological Pain

It is important to acknowledge that physical and psychological pain matter. Ultimately, the reasons we care about mortality include these things: illness and death cause suffering, physical and psychological, and if they did not, our attitudes towards them would be very different. Therefore, one cannot dismiss the fact that lockdown was painful, physically and mentally, for many people.

The pain of lockdown was strongly correlated with poverty. Overcrowding, discussed in the next section, makes staying at home exceedingly uncomfortable or even impossible at times, for example, due to high temperatures. Psychological impacts are likewise greater in less pleasant surroundings and are compounded by fears of loss of livelihood, domestic violence, food insecurity, and so forth. A significant proportion of the populations in developed countries strongly favoured lockdowns, and in one remarkable UK survey, one in five respondents indicated they favoured the situation being made permanent. This implies that, for them, if there was physical or psychological pain, it was not significant. Pain hurts. Fear or other mental states might motivate one to withstand it up to a point, but there is a point beyond which it becomes intolerable. The reason that similar results are not forthcoming from low-income settings, in any country, is in part that the pain of lockdown was greater. The will to withstand pain also dwindles over time, and more stringent, longer-lasting lockdowns resulted in protests. The role of pain cannot be set aside as a dimension along which lockdowns placed greater burdens on some than on others.

10.5 Unequal Distribution of Benefits

The benefit of lockdown in the context of the Pandemic was intended to be reduced mortality from Covid-19. Some claimed unintended benefits, such as reduced CO_2 emissions. However, there has not been any serious argument that unintended benefits justified or could justify lockdowns, and such an argument would need to overcome serious obstacles more related to retrospective justification and justification of continuing with a policy for different reasons than those that were material in its enactment. We therefore set unintended benefits aside and consider the benefit of lockdown as being the extent to which it reduced Covid-19 mortality.

10.5.1 Lockdown Is Not Equally Effective Everywhere

Slums are concentrated in low- and lower middle-income countries, and in urban areas. In its sustainable development goals, the UN estimates that 1.1 billion (or one in three) people living in urban areas live in a slum (UN Department of Economic and Social Affairs 2023). In some cities, it is much higher: for instance, in Accra (capital of Ghana), over half the population lives in a slum (Haider et al. 2020). According to the NGO Habitat for Humanity:

- 1.6 billion people live without adequate shelter.
- 1 in 7 people on the planet currently lives in a slum.
- 1 in every 4 people will live in a slum by 2030, according to current estimates.
- 1 in 3 urban residents live in slums in developing countries.
- In some countries, as much as 90% of the urban population lives in slums (Habitat for Humanity 2017).

That means that, at the time of the Pandemic, about one in seven people lived in conditions typified by:

- Inadequate access to safe water;
- Inadequate access to sanitation and other infrastructure;
- Poor structural quality of housing;
- Overcrowding; and
- Insecure residential status (UN Habitat 2007).

In these circumstances, the gap between a regulation intended to reduce social contact and the degree to which social contact is reduced becomes particularly important. Earlier, we criticized ICL12 for modelling social distancing regulations directly as reductions in contact rates because policies have different effects in different contexts. Here, the problem is that some of the intended reductions may be unachievable by any policy whatsoever. It is implausible that the highest levels of social contact reduction contemplated in modelling exercises, such as 75%, are achievable in a slum. Consequently, it is implausible that lockdown would have the same benefit in a slum as it would in non-slum conditions.

The reasons for this are twofold. First, high occupancy in the home means staying home will result in considerable social contact. This means that even if social contact outside the home does reduce significantly, Covid-19 will still spread to large numbers of people within the high-occupancy homes. It may also mean that social contact increases under lockdown conditions, because the home may be the most crowded environment that a person regularly encounters, perhaps especially in rural slums. This may explain why a possible increase in transmission of Covid-19 was observed under lockdown in some settings (Haider et al. 2020).

Second, slum living implies a need to leave the home frequently for drinking water, sanitation, and ablution, involving the use of shared facilities, if they exist. It also implies a greater need to leave the home for livelihood purposes, whether for work, or grants of food or cash. In practice, this often implies close social contact in large crowds or closely packed queues, which were sometimes kilometres long. Again, this implies that a lockdown offers reduced benefit to slum dwellers, since, even if they comply with regulations permitting only essential activities, this may still imply frequent, extended, and/or high-contact periods outside the home.

A large proportion of the global poor are employed in agriculture. In rural areas, slums may be less common, but those involved in agriculture will often experience some of the same challenges, where they are living in conditions of poverty. Rural areas in low-income settings often lack piped water. Crops and animals require constant attention and cannot be left alone during a period of lockdown. Whether these amount to legitimate essential activities or require a violation depends on the regulation in question. National food security depends on farming continuing, so agriculture is typically deemed essential, but even if it were not, the incentive to violate lockdown regulations would be overwhelming. Either way, the benefit of the lockdown order is inevitably limited by this factor in an agricultural context.

We now present two illustrative examples: one of the differential effectiveness of lockdowns in slums and suburbs, and one showing the lack of difference in Covid-19 outcomes between neighbouring nations with significant low-income populations that did and did not lock down.

10.5.1.1 Example: Townships and Suburbs in South Africa

In South Africa, the first lockdown occurred and was lifted before the first wave had peaked. The lifting of the lockdown did not have a large effect. This is not to say that it had no effect at all, simply that the benefit of the early lockdown was not a major one, which, of course, matters to the justification of lockdown. This raises a question as to why.

Part of the answer may be that lockdowns had low effectiveness among lower-income groups. This is made plausible by comparing slums and suburbs in the same area. Figure 10.1 shows the first and second epidemic waves in the townships (including slums) around the city of Cape Town, and Figure 10.2 shows the same period for suburbs in the same area.

The experience of the two waves was very different in the two environments. In the suburbs, the first wave was much smaller than the second wave. South Africa did not lock down in the second wave, or at any subsequent time. The

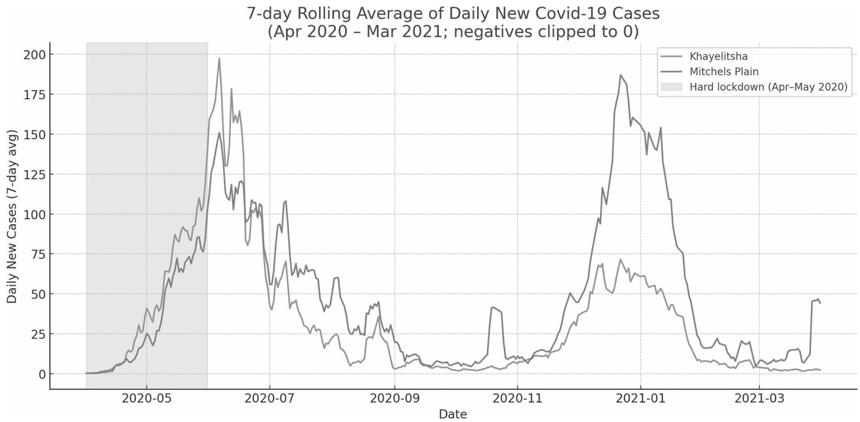


FIGURE 10.1 Mortality in townships (including slums) near Cape Town in the first two waves

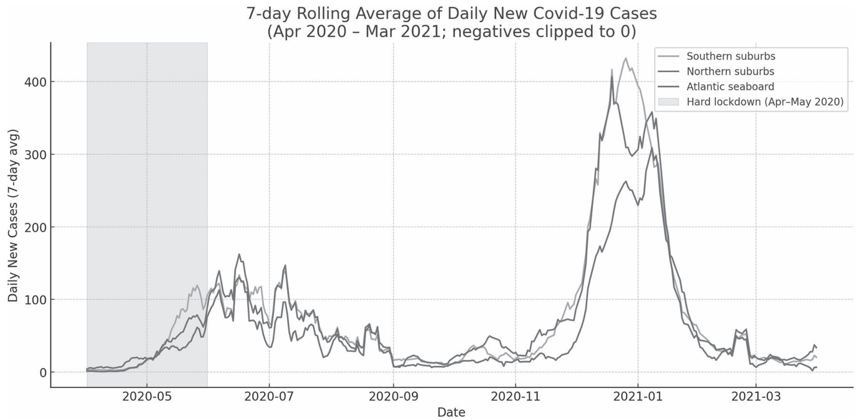


FIGURE 10.2 Mortality in suburbs near Cape Town in the first two waves

result was very significant mortality in areas that had locked down previously, a pattern everywhere due to lower levels of prior infection. In the townships, on the other hand, the two waves are approximately the same size. The most plausible explanation, indeed the only one we are aware of, is that lockdown was effective in the suburbs of Cape Town but not (at least, not very) effective in the townships.

This is an illustration; comparisons of this kind can be highly misleading. However, it is important as a test of our claim, since if the opposite were observed, this would be evidence against it. Thus, we take it as an illustration of what one would expect to find if, as we suggest, lockdowns offer lower benefits to slum dwellers in particular and poor people more generally.

10.5.1.2 Example: Peru versus Brazil

Conversely, if lockdowns have low effectiveness in many low-income settings, then one would expect low-income settings implementing opposing lockdown policies to have roughly comparable outcomes. Here, the contrast between Peru and Brazil is illustrative, with the same caveats as previously. Peru had some of the most stringent and long-lasting lockdowns, but had worse outcomes than Brazil, and indeed some of the worst in the world (Schwalb and Seas 2021). This was not widely discussed in the international media, which focused attention on the Brazilian president Jair Bolsonaro and his dramatic and implausible statements on Covid-19. Peru was initially held up in positive contrast to Brazil, as a nation taking decisive action. However, this action did not prevent Peru from having worse outcomes than Brazil. It is indicative that this did not receive significant media coverage.

International comparisons are even more fraught than comparisons within a nation or sub-national region. We therefore position this, once again, as a potentially falsifying test case for our thesis. Peru and Brazil are different, but there are important similarities: challenges with healthcare, high informal employment, inequality, high urban density, high level of co-morbidities, indigenous population at higher risk, and delayed vaccination rollouts. One part of the explanation for the fact that Peru did not have significantly better outcomes than Brazil is that lockdowns were not significantly effective in either country. This explanation has a number of merits: in particular, we have a plausible mechanistic explanation, given above; and, even if there was a benefit of lockdown in Peru, it cannot have been significant, since to assert this implies the difference would have been substantially larger if there had been no lockdown, a conclusion requiring significant speculative reasoning.

10.5.2 Lower Absolute Threat

The benefit of averting Covid-19 mortality depends on the magnitude of the threat that Covid-19 poses. This can be understood in two ways: the absolute number of lives Covid-19 puts at risk, and the threat that Covid-19 poses relative to other threats. We begin by considering absolute threat.

The UK Office of National Statistics, in response to a freedom of information inquiry, said in January 2021 that the median age of a Covid-19 death in the United Kingdom was 83, and the mean was 80. The median age of the population of sub-Saharan Africa in 2020 was 18.6 (Worldometer 2024). Mortality from Covid-19 varied, of course, but in this age range, it is tiny everywhere. Conversely, the proportion of the African population over 60 is 5.6%, compared to 23% in the United States, 25% in Europe, and double-digit figures everywhere else (He 2022).

Africa (sub-Saharan) deserves special attention here because it is by far the youngest continent, and by far the poorest. Extreme poverty is increasingly concentrated here: in 1990, 14% of the world's extreme poor (<$2.5 per day) lived in Africa; in 2019, it was 57%, and the proportion is projected to continue increasing (Schoch and Lakner 2020; Hoogeveen, Mistiaen, and Wu 2024). Africa is the second most populous, its 1.3 billion people in 2020, exceeding the total population of North America and Europe (including Russia) combined. There were and, at the time of writing, still are more children in sub-Saharan Africa than the entire population of North America. Most African countries implemented a lockdown, but it is implausible in the extreme that this offered much benefit to those children, because the threat posed by Covid-19 was so low.

It is often countered that co-morbidities associated with poverty increased the threat of Covid-19. This is true and explains higher mortality in low-income groups in countries besides low-income countries. The picture is therefore not so straightforward as saying that the poor, being young, were at less risk of Covid-19. However, *a significant proportion* of them were, and that is *the poorest proportion*. This is probably because the co-morbidities of poverty that correlate with increased Covid-19 mortality, especially obesity, are less prevalent in very low-income groups, for whom calorie deficit rather than calorie surplus is the problem. It is also probably because the very low-income groups are also the youngest.

Thus, for at least *some* of the global poor, especially the poorest, lockdown offered lower benefits due to a lower absolute threat of Covid-19 mortality. This is particularly visible in Africa, where Covid-19 mortality was notably low in many countries. Some have subsequently sought to argue that this is due to widespread under-ascertainment, backing this up with models purporting to show that excess mortality was high – some even claiming it was the highest of all regions. However, Kate Meagher argues convincingly that such "remastering" of the data shows more about the assumptions of the modellers than about the impact of Covid-19 on the continent:

> the universalizing urge inherent in global modelling rides roughshod over evidence that the trajectory of COVID-19 in Africa was different from other regions.
>
> *(Meagher 2023)*

The effort to show that Covid-19 took an enormous, undetected toll in Africa is sophisticated speculation, and its results are implausible. First, they fail sense checks, a theme to which we return in Chapter 4 when we consider predictive modelling. Even if there were more than the recorded 29,000 deaths in East Africa, could a million have gone unnoticed, as one model implies? Could 50 times as many people have died of Covid-19 as of Ebola in West Africa, as one model implies, without the local population cottoning on? The assumptions at play in supposing so are worthy of interrogation. Whatever the difficulties with recording data, death does not go unnoticed in Africa, in local communities, in families, and in society at large. Death is a major event; funerals are traditionally large. No wonder that the health minister of Sierra Leone (trained as an epidemiologist) sounded frustrated:

> We haven't had overflowing hospitals. We haven't. There is no evidence that excess deaths are occurring.
>
> *(Nolen 2022)*

Second, there are studies with more contextualized assumptions and more credible findings. For example:

> The African region is estimated to have had a similar number of COVID-19 infections to that of the rest of the world, but with fewer deaths.
>
> *(Cabore et al. 2022)*

This might not be what people expected when they predicted "bodies in the streets" (CNN 2020), but it is what the evidence suggests. Given that most of the continent is under 20, we think that the "vulnerability" of Africa ought not to have been assumed at the outset. "Vulnerable" is not an intrinsic property: it is relative to the threat. We suspect that many were, and are, so used to thinking of Africa in the victim role that they did not, and sometimes still do not, see the available information clearly.

10.5.3 Lower Relative Threat

Even where the absolute threat from Covid-19 was high, it posed a lower relative threat to lower-income groups. This is because they face other threats to life. About 5.2 million children under 5 years old are estimated to have died in 2019 (Sharrow et al. 2022). This is significantly more than Covid-19 mortality across all age groups in any given year, which peaked in 2021 at around 3.5 million (WHO 2023). The majority of these under-5 deaths occur in sub-Saharan Africa and are due to causes preventable by existing means, which is why they are rare in higher-income contexts. There is a link here to the low absolute risk: the reason that the median age in extremely poor contexts is low is that children, and people more generally, are already dying in large numbers from other serious diseases, as well as from accidents or war. Infectious disease dominates the leading causes of death in low-income countries. If one were willing to incur a very high cost to fight infectious disease in Africa, the most effective targets would be the likes of lower respiratory tract infection, diarrheal disease, neonatal complications, and nutrition. From a public health perspective, these would much better reward the huge cost incurred in seeking to control Covid-19.

10.6 Foreseeability, Avoidability, and the Unlocking Criteria

10.6.1 Unequal Distribution Was Reasonably Foreseeable

We say the distribution of benefits or burdens of a policy is *reasonably foreseeable* for a scientific authority when a reasonable scientific authority would foresee that there was a realistic prospect that an unequal distribution would flow from the policy.

This definition of "reasonable foreseeability" thus echoes the English law of negligence. As in the legal context, "foreseeable" does not mean "foreseen." It is plausible that, early in 2020, a degree of groupthink meant that some public health professionals did not foresee that lockdowns would distribute benefits and burdens unequally, or not to the degree that they did. However, if those professionals could reasonably have foreseen this, which we will argue, then the reasonable foreseeability condition is satisfied even for those unforeseen consequences.

Most of the facts discussed in the preceding two sections are common knowledge and were known well before the Pandemic. Many people live in slums, work in the informal economy, die young, and so forth. There can be no serious debate about foreseeability as it relates to those things. This is why we have been uncompromising in our discussion of the WHO China Joint Mission Report and ICL12. Both failed to take into account facts that were widely known and should have been taken into account. ICL12 could, but did not, make an attempt to evaluate the effect of its assumptions about the relevance of hospital bed capacity or the feasibility of social contact reduction in the countries where it was evaluating prospective interventions. The WHO could, but did not, have considered the ways in which China differs from every other country in the world. A reasonable scientific authority would have done so and would have foreseen the possibility of negative consequences flowing from unduly contextualized endorsement of strong social distancing policies. These consequences were therefore reasonably foreseeable.

Other facts concern Covid-19 itself, and these were sometimes objects of legitimate uncertainty. However, the points of uncertainty were relatively few. The way that Covid-19 affected people of different ages was observed already in January (Verity et al. 2020), and well-established by March, when the first African lockdowns occurred. A major uncertainty concerned how Covid-19 would affect populations with a high prevalence of HIV and TB. However, reassuring evidence on this question began emerging in March (Alcorn 2020). This did not affect the lockdown policy, since it was not the real driver of the lockdown.

There was no cogent evaluation of how lockdown might play out in light of what was known about contexts of life in much of the world, nor of whether it would have an effect on a scale that would make it worthwhile, over and above less damaging measures. Such assessments are complicated, but so is modelling infectious disease transmission. These things could have been done. The unequal distribution of benefits and burdens of lockdown was reasonably foreseeable.

10.6.2 Unequal Distribution Was Avoidable

We say that an unequal distribution of the benefits and harms of a policy is *avoidable* when viable and acceptable alternatives exist (which are distributively more equal, and epistemically accessible). Were there viable and acceptable alternatives to lockdown?

The strongest objection to the argument that lockdown was negligent or unjust is that it represented a better solution than any others *within* this acceptable range of solutions. Maybe other solutions were more equitable, and maybe some of these were even morally acceptable, the objector might concede, but lockdown still represented a better balance of the competing moral considerations of equality and reduction of Covid-19 mortality. This is an important objection.

Three interventions are much more equitable and provide most of the benefit of a full lockdown, especially in low-income settings: case isolation, protection of at-risk groups, and gathering limits. Travel restrictions are not included on this list because they did not prove particularly effective in most cases and were often implemented unfairly, for instance, when Omicron was sequenced in South Africa, and the UK blocked all flights, but Omicron had already reached the UK. African countries are expected to produce new diseases, not new knowledge.

Case isolation is a standard measure for infectious diseases. Its effectiveness is limited by the existence of asymptomatic infection and by the availability of tests. However, this does not mean it should be neglected. It is also intuitive and readily comprehensible, and coincides naturally with the need to rest.

Protection of at-risk groups is likewise an obvious and intuitive measure. Once it is understood that certain groups are at higher risk, it is natural to seek to protect them. This typically places a larger burden on those groups, but they also benefit the most. In addition, structures for caring for these groups already exist to some degree. Older people and immune-compromised people are often in need of some degree of special treatment. Wealthy people will find it easier to provide this special treatment. But many poorer people can, too. A documentary interviewed Malawian villagers early on, and they proposed moving the older population to one side of the village, and the rest to the other (Gibb 2020). Rural Malawi may be poor, but it is spacious. Overcrowded situations – slums – represent the biggest challenge for the protection of at-risk groups. Nonetheless, it is a measure that can be attempted to the extent possible, while lockdown is enforced, whether possible or not. Protection of at-risk groups is far more equitable.

Gathering limits were recognized as an effective measure to reduce Covid-19 incidence. "Super-spreader" events received widespread publicity. Gathering limits impose the most severe burdens where livelihoods depend on large gatherings, such as economies where large markets play a central role. Ghana is an example. However, this is the sort of problem to which it is possible, if not easy, to adapt. Humans are problem-solving creatures. Also, gathering limits are not perfect. They will burden Ghanian traders more than street traders or Wall Street traders. They are nonetheless considerably better than lockdowns, which, anyway, imply gathering limits, and a lot more besides.

10.6.3 The WHO*'s Criteria Imply a Never-Ending Lockdown*

The foregoing considerations pre-empt the most familiar attitudes to lockdown, which are that it was an unfortunate necessity in the absence of any other alternative, or alternatively a precaution in the absence of better information. We have shown that there were alternatives and that there was adequate information to see that lockdown was too costly (in human terms) and insufficiently effective to be universally applied as a precaution in all contexts, even if there were contexts in which that argument could be mounted.

Our final piece of evidence that these contexts were neglected comes from the *WHO*'s criteria for lifting lockdowns. These were rarely met, and for them to have been met in low-income countries would have required the development of health infrastructure, taking years.

On 13 April 2020, the *WHO* Director General announced six criteria that countries should consider before lifting lockdowns. This was in response to a discussion about the prospect of easing restrictions. The message was that this should not be rushed into. The day after the press briefing, they were released in a strategy and summarized in a guidance sheet, saying that they "outline what should be in place before transitioning away from these measures currently in place, such as physical distancing and 'shutdowns'" (*WHO* 2020b).

The six criteria were as follows:

1. Transmission of the coronavirus is under control.
2. The health system is able to detect, test, isolate, and treat every case and trace every contact.
3. The risk of outbreak hotspots is minimized in vulnerable settings like health facilities.
4. Workplaces, schools, and other essential places have preventative measures in place.
5. Measures are in place to manage the risk of importing new cases.
6. Communities are fully educated, engaged, and empowered to adjust to the new norms.

(WHO 2020b)

Meeting these criteria was foreseeably impossible for the majority of countries, and foreseeably harder for poorer countries. Controlling transmission depends on the effectiveness of lockdowns, which we have already argued was foreseeably less effective in low-income contexts. Health systems in low-income contexts are unable to offer essential treatment for existing health burdens, and would not be able to detect, test, isolate, and treat every case of a novel epidemic disease without a miraculous leap in development and resourcing. Managing the risk of importing new cases implies effective border controls, but many countries have porous borders. Educating and empowering communities implies a degree of state development that is simply absent in many places, where education provision is already highly uneven, and where empowerment is at odds with authoritarian or dismissive attitudes to significant sections of the populace. In short, for many of the places where the global poor live, the *WHO*'s criteria were not so much for lifting lockdown as for enacting sweeping reforms of the healthcare system and wider society. If they had been applied, most of the world's poor would still be in lockdown. This shows that the global poor were neglected.

10.7 Is Scientific Orthodoxy Necessarily Scientifically Unjust?

One might wonder whether scientific injustice is really a necessary part of scientific orthodoxy. Could methodological rigidity, scientific dogma, suppression of dissent, and illegitimate IPA exist without scientific injustice? Imagine a sceptic arguing as follows. So far, what has been described is essentially a set of epistemic failings. Even where the critique concerns the political reach of science, its core is epistemological: IPA is legitimate in virtue of what scientists know, and illegitimate when something goes wrong here. But scientific injustice is quite different, says the critic: it is a matter of science producing socially unjust consequences. Why could not one have methodological rigidity, scientific dogma, suppression of dissent, and illegitimate IPA without social injustice, asks the sceptical critic? Perhaps the resulting situation would still be ethically problematic – this critic need not be a staunch defender of a strong fact/value firewall, though they could be. But it is still not obvious, to our critic, why those four elements – bad though they may be – must necessarily lead to scientific injustice as we think of it here, a foreseeable, avoidable, unequal distribution of benefits and burdens. If they are feeling clever, the critic might even construct a dilemma: if the other elements do entail scientific injustice, then we do not need to include it, while if they do not, we need a reason to include it. But more plainly, the critic is simply saying: must all scientific orthodoxy, understood as a series of epistemic failings, captured in the first four elements of our account, result in the kind of injustice we have identified here?

We accept part of what the critic says: that there is nothing in the first four elements of scientific orthodoxy linking them to matters of equality. That political idea is absent from the other four elements, but present in our definition of scientific injustice. On the other hand, we reject the idea that the other four elements of scientific orthodoxy could exist with perfect justice. The core idea of our notion of scientific injustice is not equality, but negligence. If policy is made on the basis of science that is methodologically rigid and dogmatic, that suppresses dissent, and that wields illegitimate IPA, then our sense of justice would be offended. Our position, then, is that some negligence-type injustice follows from any instantiation of the other four elements of scientific orthodoxy. We have characterized it in terms of inequality on the basis of our working premise, indicated previously, that the benefits and burdens of science should be shared equally. We have adopted this working premise because (a) we suspect it is shared by most scientists involved in public health, and (b) the distribution of the benefits and burdens of Covid-19 response measures were seriously inequitable. As indicated, we are not egalitarians, but pragmatists critiquing a situation so inequitable as to disquiet anyone, egalitarian or otherwise. And in practice, we strongly suspect that any distribution of benefits and burdens on the basis of bad scientific advice will be socially unjust, unless by a remarkable chance.

Our critic could insist on a non-egalitarian version of scientific orthodoxy, one that made no reference to unequal burdens. We would be happy with that adaptation; as indicated at the outset, we intend our account to be adapted in use. However, the critic could not drop the notion of scientific injustice altogether, because the core of that notion is the idea of negligence. If a policy is implemented on the basis of bad scientific advice, the resulting distribution of benefits and burdens will be unwarranted in the epistemic sense; and, if this was foreseeable (if the advice should have been better), this means it will also be unjustified from a moral perspective, and hence unjust. That is the core reason we believe that scientific injustice is part of scientific orthodoxy, and not merely a contingent consequence. How one cashes out this idea depends on other issues concerning the relation between burdens, risks, and harms, and we cannot cover all those here, but we will do enough to show that there is a connection, and that it is strong.

For the purposes of this section, let us use SO to refer to the other four elements of scientific orthodoxy (methodological rigidity, scientific dogma, suppression of dissent, and illegitimate IPA), but not scientific injustice. SO shapes policy, as we have argued extensively in this book. This means that it affects the distribution of benefits and burdens in some way, since it is a social fact that policy affects the distribution of benefits and burdens (a policy that does not do this is a failure). If a policy is guided by SO, then its epistemic warrant is defective, due to the epistemic problems with SO that we have detailed extensively already. But when there is SO, there will be better alternative policies available: feasible policies with stronger epistemic warrant. If there were not, then there would be no need for its advocates to go to the trouble of defending SO – which is not a mere fixed mindset, but a strenuous, active engagement in the political world, sometimes coming with personal risks and costs. And a reasonable scientific authority would foresee both the defective warrant and the way this would affect the distribution of benefits and burdens. Imposing a distribution on the basis of a foreseeably defective warrant, when better-warranted options were available, is the essence of scientific injustice. It is a negligence-type wrong: the authority negligently chooses the less-optimal course and thereby wrongfully burdens those subject to the policy.

To restate the argument in a more structured way (⇒ stands for entailment):

1 Influence: SO ⇒ policy is guided by SO.
2 Distribution: any policy ⇒ some distribution of benefits/burdens.
3 Defective warrant: SO-guided policy ⇒ epistemically defective warrant.
4 Better alternatives: in SO contexts, better-warranted, feasible policies exist.
5 Foreseeability: a competent authority could foresee 3 and 4.
6 Negligence principle: choosing the SO-guided distribution despite 5 constitutes scientific injustice (SI).
7 Conclusion: scientific orthodoxy entails scientific injustice (SO ⇒ SI).

To see how this follows, we can formulate the following predicates.

Ox = SO obtains in context x
Px = Policy in x is guided by SO
Dx = A distribution results
Jx = The warrant for that distribution is defective
Bx = A better-warranted, feasible alternative exists
Fx = A competent authority could foresee J and B
Ix = Scientific injustice obtains

The premises above are then reflected by the following axioms.

1 $\forall x\,(Ox \rightarrow Px)$
2 $\forall x\,(Px \rightarrow Dx)$
3 $\forall x\,(Px \rightarrow Jx)$
4 $\forall x\,(Ox \rightarrow Bx)$
5 $\forall x\,(Ox \rightarrow Fx)$
6 $\forall x\,((Dx \wedge Jx \wedge Bx \wedge Fx) \rightarrow Ix)$

The derivation is as follows.

$Ox \Rightarrow Px \Rightarrow (Dx \wedge Jx)$ *(1–3)*
$Ox \Rightarrow (Bx \wedge Fx)$ *(4–5)*
Hence $Ox \Rightarrow (Dx \wedge Jx \wedge Bx \wedge Fx)$
Apply $6 \Rightarrow Ox \Rightarrow Ix$ *(Q.E.D.)*

This may be complicated, but it is not particularly surprising. When policies are based on defectively warranted scientific advice, despite the reasonable availability of better-warranted alternatives, they perpetrate an injustice upon those who bear the resulting burdens. Of course, the argument does not derive the full content of scientific injustice as we have defined it. But it demonstrates that SO (methodological rigidity, scientific dogma, suppression of dissent, and illegitimate IPA) is sufficient for a negligence-type injustice. That sufficiency provides a strong rationale for defining a negligence-based notion of scientific injustice and including it in the definition of scientific orthodoxy. In short, scientific orthodoxy inevitably produces negligence-type injustice, because it foreseeably and avoidably distributes benefits and burdens in sub-optimal ways.

10.8 Conclusion

It is often said that Covid-19 "laid bare pre-existing inequalities." The conclusion is then sometimes drawn that we need to mend society so as to reduce these. While we agree with the conclusion, we regard this line of response as a smokescreen, concealing a much more proximate target of evaluation: the policy choices that were made in response to Covid-19. These were tangible and could have been made differently. Unlike the project of reducing social inequality, it was eminently achievable for the scientific community, the *WHO*, and other authorities speaking with the authority of science to direct the policy of poorer countries differently.

In fact, the first reaction of these authorities was to urge a lockdown, and thus to transfer the burden of Covid-19 from rich to poor, old to young, man to woman, white to black, and so on. We are unable to think of a single less-privileged group that benefited more or suffered less from lockdown than their globally or locally more fortunate counterparts.

This is, in part, because Covid-19 was unusual in posing a much greater danger to the old than the young. The situation would have been markedly different had Covid-19 been a much larger threat to the young than the old, or posed an even greater threat across the population, or had some other pattern. However, we are not confident that the policy outcome would therefore have been fairer because we are not confident that the inequality of lockdown was not part of the reason it happened.

Lockdown in a comfortable house in a pleasant suburb, with a laptop and food delivered to your door, is a completely different prospect from lockdown in a shack with ten other people, with no food or water, and at risk of violence, molestation, or infectious disease. This fact cannot be ignored when explaining why lockdown happened. It is naïve to explain lockdown solely as the outcome of a rational process, conducted for the benefit of all. It is well-known that scientific research is strongly influenced by financial incentives. We also benefit from several decades of intensely critical analysis by sociologists of scientific knowledge, science studies scholars, historians of science, philosophers of science, and others, showing how often science serves the interests of those who do it, or of the class or other group to which they belong. The distribution of burdens and benefits of lockdowns is so strikingly unequal that it is inconceivable that such interests should be ignored. Yet they largely were, and are. Historians of the future will certainly see the role that interests played, but at present, we remain mostly blinded by proximity.

Notes

1 This chapter draws upon material in (Broadbent and Smart 2020, Broadbent and Streicher 2022, Broadbent and Streicher 2025, Smart and Broadbent 2020).

2 The report does acknowledge this in passing, but turns it into an argument in favour of greater rigour in high-income countries so as to protect the low-income countries (WHO 2020a, 20). It does not say what those lower-income countries should do, nor properly acknowledge the wide variation in relevant coping capacities among middle- and even high-income countries.

11

CONCLUSION

This book has been a sense-making exercise. Other ways of making sense are possible, and especially we expect that histories of the Pandemic will come to be written. Our approach has been from the perspective of applied philosophy of science, and our output is ultimately intended to be useful and useable. For this reason, we have provided our theory as a series of tests. The tests are therefore to be treated as the basis for development in application, as we explained in Chapter 1. This is meant to help avoid a scientific orthodoxy emerging, especially in an emergency, where the risk is high.

We take seriously Whitty's admonishment to scientists not to formulate policy (Whitty 2015), and assume it applies equally to philosophers. However, as we argued in Chapter 9, the political power of science in contemporary society requires more governance than scientific governance structures currently provide. This is both to curtail and to enable. Contemporary societies need scientific input into decisions. When it comes to structures to enable this interaction during emergency situations, there is currently a near-vacuum. We have tried to begin the task of filling the vacuum.

Who is supposed to apply these tests? Although we have left this mostly open, it is clear that they cannot be internally adjudicated within science. Generally, we would envisage a scientific advisory group considering these tests and submitting an explanation of how the corresponding risks have been dealt with when submitting their advice to ministers, or whatever is the relevant political authority in that jurisdiction. So, for example, the UK's SAGE might have included in its reports an explanation of why there was no methodological rigidity, which would have required an explanation of which methods were used and why some

DOI: 10.4324/9781003732495-12

were preferred over others. It would have had to consider whether its proposals were dogmatic, and this would have meant indicating what recalcitrant evidence there was, and how it had been dealt with. It would have considered whether there was a risk of suppression of dissent, and explained what opportunities remained for competition between ideas. It might have indicated whether its recommendations had the force of indirect political authority, and if so, why – that is, whether there was exclusive epistemic access, a shared background, and a keystone fact. Finally, it could have indicated whether its recommendations might have foreseeable, avoidable unequal distribution of burdens and benefits. Ultimately, it is a minister or other political leader who decides what to do, and whether the tests are satisfied; but the process of explaining itself relative to these tests would surely have helped the UK's SAGE and equivalent bodies in other countries to formulate better advice, and such self-evaluation would surely have helped leaders make an informed decision.

At present, there is a great deal of focus on pandemic preparedness. We hope that this book is a helpful contribution. However, we also suspect that the next Pandemic may not be a pandemic: the next time science and policy interact in an unforeseen emergency, it will of necessity be one we did not foresee. This is why it is vital to go beyond thinking about infectious respiratory diseases, vaccines, and so forth, and look for some deeper lessons. If science is to be helpful in an emergency, then we must prepare not just for pandemics in particular, but for scientific emergencies generally. That is what we hope to have helped with.

REFERENCES

Abaluck, Jason, Laura H. Kwong, Ashley Styczynski, Ashraful Haque, Md Alamgir Kabir, Ellen Bates-Jefferys, Emily Crawford, et al. 2022. “Impact of Community Masking on COVID-19: A Cluster-Randomized Trial in Bangladesh.” *Science (New York, N.Y.)* 375 (6577): eabi9069. https://doi.org/10.1126/science.abi9069.

Abbott, Sam, Joel Hellewell, Katharine Sherratt, Katelyn Gostic, Joe Hickson, Hamada S. Badr, Michael DeWitt, James M. Azam, and Sebastian Funk. 2024. “EpiNow2 – Estimate Real-Time Case Counts and Time-Varying Epidemiological Parameters.” epiForecasts. https://epiforecasts.io/EpiNow2/.

Adair, Tim. 2021. “Who Dies Where? Estimating the Percentage of Deaths That Occur at Home. Original Research.” *BMJ Global Health* 6 (9): e006766. https://doi.org/10.1136/bmjgh-2021-006766.

African Development Bank. 2024. “African Economic Outlook 2024.” African Development Bank Group. https://www.afdb.org/en/knowledge/publications/african-economic-outlook.

Alcorn, Keith. 2020. “COVID-19 in Barcelona: First Cases in People Living with HIV Reported.” *Aidsmap.Com*, April 16. https://www.aidsmap.com/news/apr-2020/covid-19-barcelona-first-cases-people-living-hiv-reported.

Alwan, Nisreen A., Rochelle Ann Burgess, Simon Ashworth, Rupert Beale, Nahid Bhadelia, Debby Bogaert, Jennifer Dowd, et al. 2020. “Scientific Consensus on the COVID-19 Pandemic: We Need to Act Now.” *The Lancet* 396 (10260): e71–72. https://doi.org/10.1016/S0140-6736(20)32153-X.

Anon. n.d. “Peer Review File: Estimating the Effects of Non-Pharmaceutical Interventions on COVID-19 in Europe.” *Nature*. Accessed August 16, 2024. https://static-content.springer.com/esm/art%3A10.1038%2Fs41586-020-2405-7/MediaObjects/41586_2020_2405_MOESM3_ESM.pdf.

Arbel Ronen, Hammerman Ariel, Sergienko Ruslan, Michael Friger, Alon Peretz, Doron Netzer, and Shlomit Yaron. 2021. “BNT162b2 Vaccine Booster and Mortality Due to Covid-19.” *New England Journal of Medicine* 385 (26): 2413–20. https://doi.org/10.1056/NEJMoa2115624.

Arega, Balew, Abebe Negesso, Betelhem Taye, Getachew Weldeyohhans, Bekure Bewket, Tesfaye Negussie, Ayele Teshome, and Getabalew Endazenew. 2022. “Impact of COVID-19 Pandemic on TB Prevention and Care in Addis Ababa, Ethiopia: A Retrospective Database Study." Infectious Diseases. *BMJ Open* 12 (2): e053290. https://doi.org/10.1136/bmjopen-2021-053290.

Arundel, Thomas. 1408. "Archbishop Thomas Arundel's Constitutions against the Lollards." Bible Researcher. https://www.bible-researcher.com/arundel.html.

Aschwanden, Christie. 2021. "Five Reasons Why COVID Herd Immunity Is Probably Impossible." *Nature* 591 (7851): 520–22. https://doi.org/10.1038/d41586-021-00728-2.

Baker, Michael G., David Durrheim, Li Yang Hsu, and Nick Wilson. 2023. "COVID-19 and Other Pandemics Require a Coherent Response Strategy." *The Lancet* 401 (10373): 265–66. https://doi.org/10.1016/S0140-6736(22)02489–8.

Basshuysen, Philippe van, and Lucie White. 2021. "Were Lockdowns Justified? A Return to the Facts and Evidence." *Kennedy Institute of Ethics Journal* 31 (4): 405–28.

Begley, Sharon. 2020. "Influential Covid-19 Model Uses Flawed Methods and Shouldn't Guide U.S. Policies, Critics Say." *Statnews*, April 17. https://www.statnews.com/2020/04/17/influential-covid-19-model-uses-flawed-methods-shouldnt-guide-policies-critics-say/.

Berger, Elke, Juliane Winkelmann, Helene Eckhardt, Ulrike Nimptsch, Dimitra Panteli, Christoph Reichebner, Tanja Rombey, and Reinhard Busse. 2022. "A Country-Level Analysis Comparing Hospital Capacity and Utilisation during the First COVID-19 Wave across Europe." *Health Policy*, Lessons Learned from the COVID-19 Pandemic 126 (5): 373–81. https://doi.org/10.1016/j.healthpol.2021.11.009.

Birch, Jonathan. 2021. "Science and Policy in Extremis: The UK's Initial Response to COVID-19." *European Journal for Philosophy of Science* 11 (3): 90. https://doi.org/10.1007/s13194-021-00407-z.

Bloor, David. 1991. *Knowledge and Social Imagery*. 2nd ed. University of Chicago Press.

Bloor, David. 2007. "Epistemic Grace: Anti-Relativism as Theology in Disguise." *Common Knowledge* 13: 250–80.

Blumenthal-Barby, J.S. 2012. "Between Reason and Coercion: Ethically Permissible Influence in Health Care and Health Policy Contexts." *Kennedy Institute of Ethics Journal* 22 (4): 345–66.

Bollyky, Thomas J., Emma Castro, Aleksandr Y. Aravkin, Kayleigh Bhangdia, Jeremy Dalos, Erin N. Hulland, Samantha Kiernan, et al. 2023. "Assessing COVID-19 Pandemic Policies and Behaviours and Their Economic and Educational Trade-Offs across US States from Jan 1, 2020, to July 31, 2022: An Observational Analysis." *The Lancet* 401 (10385): 1341–60. https://doi.org/10.1016/S0140-6736(23)00461–0.

Boulos, Leah, Janet A. Curran, Allyson Gallant, Helen Wong, Catherine Johnson, Alannah Delahunty-Pike, Lynora Saxinger, et al. 2023. "Effectiveness of Face Masks for Reducing Transmission of SARS-CoV-2: A Rapid Systematic Review." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 381 (2257): 20230133. https://doi.org/10.1098/rsta.2023.0133.

Boyd, Lydia. 2023. "Elections, COVID-19 Lockdowns, and the Politics of Health in Uganda." *Africana Public Interest Journal* 1: 52–57.

Braver Angels, dir. 2023. "A Deplorable and an Elitist Walk into a Bar: Francis Collins and Wilk Wilkinson." Braver Angels. https://www.youtube.com/watch?v=W1eAvh1sWiw.

Brister, Evelyn, and Robert Frodeman. 2020. *A Guide to Field Philosophy: Case Studies and Practical Strategies*. Routledge.

Britton, Tom, Frank Ball, and Pieter Trapman. 2020. "A Mathematical Model Reveals the Influence of Population Heterogeneity on Herd Immunity to SARS-CoV-2." *Science* 369 (6505): 846–49. https://doi.org/10.1126/science.abc6810.

Broadbent, Alex. 2013. *Philosophy of Epidemiology*. Palgrave Macmillan.

Broadbent, Alex. 2021. "The C-Word, the P-Word, and Realism in Epidemiology." *Synthese* 198 (Suppl 10) (May): 2613–2628. https://doi.org/10.1007/s11229-019-02169-x.

Broadbent, Alex. 2020. "A Framework for Decisions in a Post-COVID World." no. May. University of Johannesburg. https://philosepi.files.wordpress.com/2020/05/a-framework-for-decisions-in-a-post-covid-world-south-africa-report−1.0.pdf.

Broadbent, Alex, Herkulaas Combrink, and Benjamin T.H. Smart. 2020. "Covid-19 in South Africa." *Global Epidemiology* 2 (November): 100034. https://10.1016/j.gloepi.2020.100034.

Broadbent, Alex, and Benjamin T.H. Smart. 2020. "Why a One-Size-Fits-All Approach to COVID-19 Could Have Lethal Consequences." *The Conversation*, March 23. http://theconversation.com/why-a-one-size-fits-all-approach-to-covid-19-could-have-lethal-consequences−134252.

Broadbent, Alex, and Pieter Streicher. 2022. "Can you lock down in a slum? And who would benefit if you tried? Difficult questions about epidemiology's commitment to global health inequalities during Covid-19." *Global Epidemiology* 4 (December): 100074.

Broadbent, Alex, and Pieter Streicher. 2025. "Was Lockdown Racist?" *Ergo: An Open Access Journal of Philosophy* 12: 20. https://doi.org/10.3998/ergo.7431.

Broadbent, Alex, Jan P. Vandenbroucke, and Neil Pearce. 2016. "Response: Formalism or Pluralism? A Reply to Commentaries on 'Causality and Causal Inference in Epidemiology'." *International Journal of Epidemiology* 45 (6): 1841–51.

Bujakera, Stanis, and Ayenat Mersie. 2020. "In Parts of Africa, Police Are Accused of Excess Force amid Coronavirus Lockdowns." World. *Reuters*, April 10. https://www.reuters.com/article/world/in-parts-of-africa-police-are-accused-of-excess-force-amid-coronavirus-lockdown-idUSKCN21S0M9/.

Bundgaard, Henning, Johan Skov Bundgaard, Daniel Emil Tadeusz Raaschou-Pedersen, Christian von Buchwald, Tobias Todsen, Jakob Boesgaard Norsk, Mia M. Pries-Heje, et al. 2021. "Effectiveness of Adding a Mask Recommendation to Other Public Health Measures to Prevent SARS-CoV-2 Infection in Danish Mask Wearers." *Annals of Internal Medicine* 174 (3): 335–43. https://doi.org/10.7326/M20-6817.

Bwire, Thomas. 2023. "Why Large Dependence on Foreign Aid yet Only Small Detectable Growth Effects." COMESA Monetary Institute. https://www.comesa.int/wp-content/uploads/2023/03/230316_Why-large-dependence-on-foreign-aid-yet-only-small-detectable-growth-effects.pdf.

Cabore, Joseph Waogodo, Humphrey Cyprian Karamagi, Hillary Kipchumba Kipruto, Joseph Kyalo Mungatu, James Avoka Asamani, Benson Droti, Regina Titi-Ofei, et al. 2022. "COVID-19 in the 47 Countries of the WHO African Region: A Modelling Analysis of Past Trends and Future Patterns." *The Lancet Global Health* 10 (8): e1099–114. https://doi.org/10.1016/S2214-109X(22)00233–9.

Camera, Gabriele, and Alessandro Gioffré. 2021. "The Economic Impact of Lockdowns: A Theoretical Assessment." *Journal of Mathematical Economics* 97 (December): 102552. https://doi.org/10.1016/j.jmateco.2021.102552.

Campbell, Charlie. 2020. "South Korea's Health Minister on How His Country Is Beating Coronavirus Without a Lockdown." *TIME*, May 1. https://time.com/5830594/south-korea-covid19-coronavirus/.

Campos-Rudinsky, Thana C. de, Sarah L. Bosha, Daniel Wainstock, Sharifah Sekalala, Sridhar Venkatapuram, and Caesar Alimsinya Atuire. 2024. "Decolonising Global Health: Why the New Pandemic Agreement Should Have Included the Principle of Subsidiarity." *The Lancet Global Health* 12 (7): e1200–3. https://doi.org/10.1016/S2214-109X(24)00186–4.

Cartwright, Nancy. 2010. "Presidential Address Will This Policy Work for You? Predicting Effectiveness Better: How Philosophy Helps." *Philosophy of Science* 79 (5): 973–89. https://doi.org/10.1086/668041.

Cartwright, Nancy, and Jeremy Hardie. 2012. *Evidence Based Policy: A Practical Guide to Doing It Better*. Oxford University Press.

Cash-Goldwasser, Shama, Arthur L. Reingold, Stephen P. Luby, Lisa A. Jackson, and Thomas R. Frieden. 2023. "Masks During Pandemics Caused by Respiratory Pathogens – Evidence and Implications for Action." *JAMA Network Open* 6 (10): e2339443. https://doi.org/10.1001/jamanetworkopen.2023.39443.

CDC. 2020. "Safety of Covid-19 Vaccines." Centers for Disease Control and Prevention, February 11. https://www.cdc.gov/coronavirus/2019-ncov/vaccines/safety/safety-of-vaccines.html.

CDC. 2023. "Vaccines for COVID-19 (Update of July 2023)." Centers for Disease Control and Prevention, July 13. https://www.cdc.gov/coronavirus/2019-ncov/vaccines/index.html.

CDC. 2024a. "Vaccines for COVID-19 (Update of April 2024)." Centers for Disease Control and Prevention, April 23. https://www.cdc.gov/coronavirus/2019-ncov/vaccines/index.html.

CDC. 2024b. "Stay Up to Date with COVID-19 Vaccines." Centers for Disease Control and Prevention, May 14. https://www.cdc.gov/coronavirus/2019-ncov/vaccines/stay-up-to-date.html.

CEPI. 2017. "CEPI Officially Launched." Coalition for Epidemic Preparedness Innovations, January 18. https://cepi.net//cepi-officially-launched.

Chatterjee, Srijan, Manojit Bhattacharya, Sagnik Nag, Kuldeep Dhama, and Chiranjib Chakraborty. 2023. "A Detailed Overview of SARS-CoV-2 Omicron: Its Sub-Variants, Mutations and Pathophysiology, Clinical Characteristics, Immunological Landscape, Immune Escape, and Therapies." *Viruses* 15 (1): 1. https://doi.org/10.3390/v15010167.

Chikina, Maria, Wesley Pegden, and Benjamin Recht. 2022. "Re-Analysis on the Statistical Sampling Biases of a Mask Promotion Trial in Bangladesh: A Statistical Replication." *Trials* 23 (1): 786. https://doi.org/10.1186/s13063-022-06704-z.

Childress, James F., Ruth R. Faden, Ruth D. Gaare, Lawrence O. Gostin, Jeffrey Kahn, Richard J. Bonnie, Nancy E. Kass, Anna C. Mastroianni, Jonathan D. Moreno, and Phillip Nieburg. 2002. "Public Health Ethics: Mapping the Terrain." *The Journal of Law, Medicine & Ethics: A Journal of the American Society of Law, Medicine & Ethics* 30 (2): 170–78. https://doi.org/10.1111/j.1748-720x.2002.tb00384.x.

Chin, John J., and Jessica Kirkpatrick. 2023. "African Coups in the COVID-19 Era: A Current History." *Frontiers in Political Science* 5: 1077945. https://doi.org/10.3389/fpos.2023.1077945.

Chmielewska, Barbara, Imogen Barratt, Rosemary Townsend, Erkan Kalafat, Jan Van Der Meulen, Ipek Gurol-Urganci, Pat O'Brien, et al. 2021. "Effects of the COVID-19 Pandemic on Maternal and Perinatal Outcomes: A Systematic Review and Meta-Analysis." *The Lancet Global Health* 9 (6): e759–72. https://doi.org/10.1016/S2214-109X(21)00079–6/ATTACHMENT/FEFEA3E3–948E-4ED8-B43E-1B36DE9CA089/MMC1.PDF.

Christensen, Zach. 2023. "Economic Poverty Trends: Global, Regional and National (Updated February 2023)." *Development Initiatives*, February. https://devinit.org/resources/poverty-trends-global-regional-and-national/.

CNN. 2020. "Melinda Gates: Covid-19 Will Be Horrible in the Developing World." *CNN Business*, April 10. https://edition.cnn.com/videos/business/2020/04/10/melinda-gates-coronavirus.cnn-business.

DH Pandemic Influenza Preparedness Team. 2011. "UK Influenza Pandemic Preparedness Strategy 2011." UK Department of Health, Social Services, and Public Safety. https://assets.publishing.service.gov.uk/media/5a7c4767e5274a2041cf2ee3/dh_131040.pdf.

Douglas, Heather. 2009. *Science, Policy, and the Value-Free Ideal*. University of Pittsburgh Press.

Elgin, Catherine Z. 1999. *Considered Judgment*. Princeton University Press.
Elliott, Kevin C., and Daniel J. McKaughan. 2014. "Nonepistemic Values and the Multiple Goals of Science." *Philosophy of Science* 81 (1): 1–21. https://doi.org/10.1086/674345.
Ellis-Petersen, Hannah, and Manoj Chaurasia. 2020. "India Racked by Greatest Exodus since Partition Due to Coronavirus." World News. *The Guardian*, March 30. https://www.theguardian.com/world/2020/mar/30/india-wracked-by-greatest-exodus-since-partition-due-to-coronavirus.
European Medicines Agency. 2021. "AstraZeneca's COVID-19 Vaccine: EMA Finds Possible Link to Very Rare Cases of Unusual Blood Clots with Low Blood Platelets | European Medicines Agency." April 7. https://www.ema.europa.eu/en/news/astrazenecas-covid-19-vaccine-ema-finds-possible-link-very-rare-cases-unusual-blood-clots-low-blood-platelets.
Evans, Gillian R. 2005. *John Wyclif: Myth & Reality*. Oxford University Press.
Fagen-Ulmschneider, Wade. 2024. "An Interactive Visualization of COVID-19." 91-DIVOC. https://91-divoc.com/pages/covid-visualization/.
Ferguson, Neil M. 2021. "BBC: Most Sundays." Interview by Andrew Marr. July 18. Television. https://twitter.com/BBCPolitics/status/1416681412795764739.
Ferguson, Neil M., D. Laydon, G. Nedjati Gilani, Natsuko Imai, Kylie Ainslie, Marc Baguelin, Sangeeta Bhatia, et al. 2020. "Report 9: Impact of Non-Pharmaceutical Interventions (NPIs) to Reduce COVID19 Mortality and Healthcare Demand." With Medical Research Council (MRC). Imperial College London. https://doi.org/10.25561/77482.
Findl, Johannes, and Javier Suárez. 2021. "Descriptive Understanding and Prediction in COVID-19 Modelling." *History and Philosophy of the Life Sciences* 43 (4): 1–31. https://doi.org/10.1007/s40656-021-00461-z.
Flanigan, Jessica. 2014. "A Defense of Compulsory Vaccination." *HEC Forum* 26 (1): 5–25. https://doi.org/10.1007/s10730-013-9221-5.
Flaxman, Seth, Swapnil Mishra, Axel Gandy, H. Juliette T. Unwin, Thomas A. Mellan, Helen Coupland, Charles Whittaker, et al. 2020a. "Estimating the Effects of Non-Pharmaceutical Interventions on COVID-19 in Europe." *Nature* 584 (7820): 257–61. https://doi.org/10.1038/s41586-020-2405-7.
Flaxman, Seth, Swapnil Mishra, James Scott, Neil M. Ferguson, Axel Gandy, and Samir Bhatt. 2020b. "Reply to: The Effect of Interventions on COVID-19." *Nature* 588 (7839): 7839. https://doi.org/10.1038/s41586-020-3026-x.
Friedman, Joseph, Patrick Liu, Christopher E. Troeger, Austin Carter, Robert C. Reiner, Jr., Ryan M. Barber, James Collins, et al. 2021. "Predictive Performance of International COVID-19 Mortality Forecasting Models." *Nature Communications* 12 (1): 1. https://doi.org/10.1038/s41467-021-22457-w.
Frigg, Roman. 2022. *Models and Theories: A Philosophical Inquiry*. Taylor & Francis.
Frigg, Roman, and Stephan Hartmann. 2025. "Models in Science." In *The Stanford Encyclopedia of Philosophy*, Summer 2025, edited by Edward N. Zalta and Uri Nodelman. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/sum2025/entries/models-science/.
Fuentes, Agustín. 2024a. "Statements by Scientific Organizations Can, and Should, Shape Society." *Science* 383 (6688): eado7084. https://doi.org/10.1126/science.ado7084.
Fuentes, Agustín. 2024b. "Scientists as Political Advocates." *Science* 386 (6724): eadt7194. https://doi.org/10.1126/science.adt7194.
Fuller, Jonathan. 2020. "Models v. Evidence." *Boston Review*, May 5. https://www.bostonreview.net/articles/jonathan-fuller-models-v-evidence/.
Fuller, Jonathan. 2021. "What Are the COVID-19 Models Modeling (Philosophically Speaking)?" *History and Philosophy of the Life Sciences* 43 (2): 47. https://doi.org/10.1007/s40656-021-00407-5.

Gallagher, James. 2020. “Coronavirus: Lockdowns in Europe Saved Millions of Lives.” *BBC News*, June 8. https://www.bbc.com/news/health−52968523.

Gibb, Tom. 2020. “Covid on the Breadline.” *Picturing Health*. https://www.picturinghealth.org/covid-on-the-breadline/

Godfrey-Smith, Peter. 2006. “The Strategy of Model-Based Science.” *Biology and Philosophy* 21 (5): 725–40. https://doi.org/10.1007/s10539-006-9054-6.

Goldenberg, Maya J. 2021. *Vaccine Hesitancy: Public Trust, Expertise, and the War on Science*. University of Pittsburgh Press.

Gomes, M. Gabriela M., Marcelo U. Ferreira, Rodrigo M. Corder, Jessica G. King, Caetano Souto-Maior, Carlos Penha-Gonçalves, Guilherme Gonçalves, Maria Chikina, Wesley Pegden, and Ricardo Aguas. 2022. “Individual Variation in Susceptibility or Exposure to SARS-CoV-2 Lowers the Herd Immunity Threshold.” *Journal of Theoretical Biology* 540 (May): 111063. https://doi.org/10.1016/j.jtbi.2022.111063.

Google Trends. 2024. “Google Trends: ‘Misinformation’ 17 Aug 2014–16 Aug 2024.” *Google Trends*, August 16. https://trends.google.com/trends/explore?date=2014-08-17%20 2024-08-16&q=misinformation&hl=en-GB.

Götmark, Frank, and Malte Andersson. 2020. “Human Fertility in Relation to Education, Economy, Religion, Contraception, and Family Planning Programs.” *BMC Public Health* 20 (1): 265. https://doi.org/10.1186/s12889-020-8331-7.

Grove, Jack. 2020. “Sunetra Gupta: Social Media Attacks ‘Stifling Coronavirus Debate’.” *Times Higher Education (THE)*, September 16. https://www.timeshighereducation.com/news/sunetra-gupta-social-media-attacks-stifling-coronavirus-debate.

Grüne-Yanoff, Till. 2021. “KISSing in the Time of COVID-19: Some Lessons for Model Choice.” *Argumenta* 7 (1): 39–55. https://doi.org/10.14275/2465-2334/202113.GRU.

Gupta, Sunetra. 2020. “Professor Sunetra Gupta Reveals How Crisis Has Been Ruthlessly Weaponised.” *The Daily Mail Online*, October 30. https://www.dailymail.co.uk/debate/article-8899277/Professor-Sunetra-Gupta-reveals-crisis-ruthlessly-weaponised.html.

Habitat for Humanity. 2017. “What Is a Slum? Definition of a Global Housing Crisis.” Habitat for Humanity GB. https://www.habitatforhumanity.org.uk/what-we-do/slum-rehabilitation/what-is-a-slum/.

Haider, Najmul, Abdinasir Yusuf Osman, Audrey Gadzekpo, George O. Akipede, Danny Asogun, Rashid Ansumana, Richard John Lessells, et al. 2020. “Lockdown Measures in Response to COVID-19 in Nine Sub-Saharan African Countries.” *BMJ Global Health* 5 (10): e003319. https://doi.org/10.1136/bmjgh-2020-003319.

Hallett, Baroness. 2024. “Module 1 Report – The Resilience and Preparedness of the United Kingdom UK Covid-19 Inquiry Archives.” House of Commons. https://covid19.public-inquiry.uk/documents/module-1-full-report/.

Hansard. 2022. “Covid-19: Forecasting and Modelling.” Hansard – UK Parliament, January 18. https://hansard.parliament.uk/commons/2022-01-18/debates/AB251DCA-8088-485C-BF49-3999C4EE9AC5/Covid-19ForecastingAndModelling.

Harrisberg, Kim. 2020. “South Africans Protest Police Brutality against Poor under Lockdown.” World. *Reuters*, June 9. https://www.reuters.com/article/world/south-africans-protest-police-brutality-against-poor-under-lockdown-idUSKBN23G2QP/.

He, Wan. 2022. “Increases in Africa’s Older Population Will Outstrip Growth in Any Other World Region.” *US Census Bureau*, April 13. https://www.census.gov/library/stories/2022/04/why-study-aging-in-africa-region-with-worlds-youngest-population.html.

Heinen, Alex. 2022. “Tomas Pueyo: Wie Ein Berater Aus Dem Silicon Valley Zum Pandemie-Flüsterer Wurde.” February 6. https://www.berliner-zeitung.de/open-source/tomas-pueyo-wie-ein-berater-aus-dem-silicon-valley-zum-pandemie-fluesterer-wurde-li.209302.

Hellewell, Joel. 2021. "Is COVID-19 Forecasting Bad, or Are You Just Projecting?". *Joel Hellewell's Personal Blog*, November 16. https://jhellewell14.github.io/2021/11/16/forecasting-projecting.html.

Hickok, Kimberley. 2020. "New York Has Been Crushed by COVID−19. Will Other US Cities Be Spared?" Text.Article. *LiveScience*, *Fox News*, April 6. https://www.foxnews.com/science/new-york-has-been-crushed-by-covid-19-will-other-us-cities-be-spared.

Høeg, Tracy B., Ram Duriseti, and Vinay Prasad. 2023. "Potential "Healthy Vaccinee Bias" in a Study of BNT162b2 Vaccine against Covid-19." *New England Journal of Medicine* 389 (3): 284–86. https://doi.org/10.1056/NEJMc2306683.

Holman, Bennett, and Doohyun Sung. 2025. "Evidence in Medicine." In *Oxford Handbook of Philosophy of Medicine*. Oxford University Press.

Holmberg, Christine, Stuart Blume, and Paul Greenough. 2017. *The Politics of Vaccination: A Global History*. Manchester University Press.

Hoogeveen, Johannes, Johan A. Mistiaen, and Haoyu Wu. 2024. "Accelerating Poverty Reduction in Sub-Saharan Africa Requires Stability." *World Bank Blogs*, March 5. https://blogs.worldbank.org/en/africacan/accelerating-poverty-reduction-sub-saharan-africa-requires-stability.

Howard, Jacqueline. 2021. "Pfizer Data Suggest Third Dose of Covid-19 Vaccine 'Strongly' Boosts Protection against Delta Variant." *CNN*, July 28. https://www.cnn.com/2021/07/28/health/pfizer-third-dose-data-bn/index.html.

Howick, Jeremy. 2011. *The Philosophy of Evidence-Based Medicine*. Wiley-Blackwell.

Hudson, Anne. 1988. *The Premature Reformation: Wycliffite Texts and Lollard History*. Oxford University Press.

IHME. 2020. "New COVID-19 Forecasts: US Hospitals Could Be Overwhelmed in the Second Week of April by Demand for ICU Beds, and US Deaths Could Total 81,000 by July | Institute for Health Metrics and Evaluation." Institute for Health Metrics and Evaluation, March 26. https://www.healthdata.org/news-events/newsroom/news-releases/new-covid-19-forecasts-us-hospitals-could-be-overwhelmed-second.

IHME. 2024. "About Us." Institute for Health Metrics and Evaluation. https://www.healthdata.org/about.

ILO. 2023. "Women and Men in the Informal Economy: A Statistical Update | International Labour Organization." International Labour Office. https://www.ilo.org/publications/women-and-men-informal-economy-statistical-update.

Imperial College Covid-19 Response Team. 2020. "COVID-19 Reports." Imperial College London – MRC Centre for Global Infectious Disease Analysis. https://www.imperial.ac.uk/medicine/departments/school-public-health/infectious-disease-epidemiology/mrc-global-infectious-disease-analysis/disease-areas/covid-19/covid-19-reports/.

Inglesby, Thomas V., Jennifer B. Nuzzo, Tara O'Toole, and D.A. Henderson. 2006. "Disease Mitigation Measures in the Control of Pandemic Influenza." *Biosecurity and Bioterrorism: Biodefense Strategy, Practice, and Science* 4 (4): 366–75. https://doi.org/10.1089/bsp.2006.4.366.

Investigations Team, and Robert Mendick. 2024. "AstraZeneca Admits Its Covid Vaccine Can Cause Rare Side Effect in Court Documents for First Time." *The Telegraph*, April 28. https://www.telegraph.co.uk/news/2024/04/28/astrazeneca-admits-covid-vaccine-causes-rare-side-effect/.

Jefferson, Tom, Liz Dooley, Eliana Ferroni, Lubna A. Al-Ansary, Mieke L. van Driel, Ghada A. Bawazeer, Mark A. Jones, et al. 2023. "Physical Interventions to Interrupt or Reduce the Spread of Respiratory Viruses." *Cochrane Database of Systematic Reviews* (1). https://doi.org/10.1002/14651858.CD006207.pub6.

Jewell, Nicholas P., Joseph A. Lewnard, and Britta L. Jewell. 2020. "Caution Warranted: Using the Institute for Health Metrics and Evaluation Model for Predicting the Course of the COVID-19 Pandemic." *Annals of Internal Medicine* (April 14): M20-1565. https://doi.org/10.7326/M20-1565.

John, Stephen. 2022. "Are Mandatory Vaccines Justified?" *IAI TV – Changing How the World Thinks*, February 1. https://iai.tv/articles/are-mandatory-vaccines-justified-auid−2039.
JSM. 2021. "The John Snow Memorandum." https://www.johnsnowmemo.com/.
Kahneman, Daniel. 2011. *Thinking, Fast and Slow*. Farrar, Straus and Giroux.
Kepp, Kasper P., Preben Aavitsland, Marcel Ballin, Francois Balloux, Stefan Baral, Kevin Bardosh, Howard Bauchner, et al. 2024. "Panel Stacking Is a Threat to Consensus Statement Validity." *Journal of Clinical Epidemiology* (June 17): 111428. https://doi.org/10.1016/j.jclinepi.2024.111428.
Kitcher, Philip. 2001. *Science, Truth and Democracy*. OUP.
Knuuttila, Tarja, Natalia Carrillo, and Rami Koskinen. 2024. *The Routledge Handbook of Philosophy of Scientific Modeling*. Taylor & Francis.
Krug, Allison, Josh Stevenson, and Tracy Beth Høeg. 2022. "BNT162b2 Vaccine-Associated Myo/Pericarditis in Adolescents: A Stratified Risk-Benefit Analysis." *European Journal of Clinical Investigation* 52 (5): e13759. https://doi.org/10.1111/eci.13759.
Kucharski, A. 2020. *The Rules of Contagion: Why Things Spread – and Why They Stop*. Profile Books Limited. https://books.google.it/books?id=B2c8zAEACAAJ.
Kuhn, Thomas S. 1962. *The Structure of Scientific Revolutions*. The University of Chicago Press.
Kuhn, Thomas S. 1977. *The Essential Tension: Selected Studies in Scientific Tradition and Change*. University of Chicago Press.
Kulldorff, Martin, Sunetra Gupta, and Jay Bhattacharya. 2020. "Great Barrington Declaration and Petition." Great Barrington Declaration, October 4. https://gbdeclaration.org/.
Ladhani, Shamez N. 2022. "COVID-19 Vaccination for Children Aged 5–11 Years." *Lancet (London, England)* 400 (10346): 74–76. https://doi.org/10.1016/S0140-6736(22)01245–4.
Law, Duncan. 2020. "John Edmunds on Herd Immunity – Full Transcript." June 11. https://duncanlaw.wordpress.com/2020/06/11/john-edmunds-on-herd-immunity-full-transcript/.
Lazarus, Jeffrey V., Diana Romero, Christopher J. Kopka, Salim Abdool Karim, Laith J. Abu-Raddad, Gisele Almeida, Ricardo Baptista-Leite, et al. 2022. "A Multinational Delphi Consensus to End the COVID-19 Public Health Threat." *Nature* 611 (7935): 332–45. https://doi.org/10.1038/s41586-022-05398-2.
Levy, Ian Christopher. 2002. "John Wyclif: Scriptural Logic, Real Presence, and the Parameters of Orthodoxy." *Church History* 71 (1): 27–56.
Lewis, David. 1996. "Elusive Knowledge." *Australasian Journal of Philosophy* 74 (4): 549–67.
Lexchin, J., L.A. Bero, and B. Djulbegovic. 2003. "Pharmaceutical Industry Sponsorship and Research Outcome and Quality: Systematic Review." *British Medical Journal* 326: 1167–70.
Lipton, Peter. 2004. *Inference to the Best Explanation*. 2nd ed. Taylor & Francis.
Liu, Ying, and Joacim Rocklöv. 2022. "The effective reproductive number of the Omicron variant of SARS-CoV-2 is several times relative to Delta." *Journal of Travel Medicine* 29 (3): taac037. https://doi.org/10.1093/jtm/taac037.
Longino, Helen E. 1990. *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. Princeton University Press.
Maziarz, Mariusz, and Martin Zach. 2021. "Assessing the Quality of Evidence from Epidemiological Agent-Based Models for the COVID-19 Pandemic." *History and Philosophy of the Life Sciences* 43 (1): 1–4. https://doi.org/10.1007/s40656-020-00357-4.
McGrath, Alister E. 2023. *Heresy: A History of Defending the Truth*. Harper Collins.
Meagher, Kate. 2023. "Africa's COVID-19 Statistics Highlight Bias in Excess Death Modelling." *Impact of Social Sciences*, May 11. https://blogs.lse.ac.uk/impactofsocialsciences/2023/05/11/africas-covid-19-statistics-highlight-bias-in-excess-death-modelling/.

Medić, Snežana, Cleo Anastassopoulou, Zagorka Lozanov-Crvenković, Vladimir Vuković, Nataša Dragnić, Vladimir Petrović, Mioljub Ristić, et al. 2022. “Risk and Severity of SARS-CoV-2 Reinfections during 2020–2022 in Vojvodina, Serbia: A Population-Level Observational Study.” *The Lancet Regional Health – Europe* 20 (September). https://doi.org/10.1016/j.lanepe.2022.100453.

Meyer-Rath, Gesine, Rachel A. Hounsell, Juliet R.C. Pulliam, Lise Jamieson, Brooke E. Nichols, Harry Moultrie, and Sheetal P. Silal. 2023. “The Role of Modelling and Analytics in South African COVID-19 Planning and Budgeting.” *PLoS Global Public Health* 3 (7): e0001063. https://doi.org/10.1371/journal.pgph.0001063.

Mishra, Swapnil, James A. Scott, Daniel J. Laydon, Seth Flaxman, Axel Gandy, Thomas A. Mellan, H. Juliette T. Unwin, et al. 2021. “Comparing the Responses of the UK, Sweden and Denmark to COVID-19 Using Counterfactual Modelling.” *Nature Scientific Reports* 11 (1): 1. https://doi.org/10.1038/s41598-021-95699-9.

Mistur, Evan M., John Wagner Givens, and Daniel C. Matisoff. 2023. “Contagious COVID-19 Policies: Policy Diffusion during Times of Crisis.” *Review of Policy Research* 40 (1): 36–62. https://doi.org/10.1111/ropr.12487.

Morabia, Alfredo. 2004. *History of Epidemiologic Methods and Concepts*. Birkhauser Verlag.

Munro, Alasdair P.S., and Saul N. Faust. 2020. “Children Are Not COVID-19 Super Spreaders: Time to Go Back to School.” *Archives of Disease in Childhood* (May): archdischild-2020–319474. https://doi.org/10.1136/archdischild-2020-319474.

Murray, Christopher J.L. 2020. “Forecasting COVID-19 Impact on Hospital Bed-Days, ICU-Days, Ventilator Days and Deaths by US State in the Next 4 Months | Institute for Health Metrics and Evaluation.” Institute for Health Metrics and Evaluation, March 26. https://www.healthdata.org/research-analysis/library/forecasting-covid-19-impact-hospital-bed-days-icu-days-ventilator-days.

Nabarro, David. 2020. “Interview with David Nabarro on Spectator TV.” no. October 9. Zoom. https://twitter.com/spectator/status/1314573157827858434?ref_src=twsrc%5Etfw%7Ctwcamp%5Etweetembed%7Ctwterm%5E1314573157827858434%7Ctwgr%5E8ad8f69c577f19833ec81337abe9abf0bbd1b31a%7Ctwcon%5Es1_c10&ref_url=https%3A%2F%2Fwww.abc.net.au%2Fnews%2F2020-10-12%2Fworld-health-organization-coronavirus-lockdown-advice%2F12753688.

Nasa, Prashant, Ravi Jain, and Deven Juneja. 2021. “Delphi Methodology in Healthcare Research: How to Decide Its Appropriateness.” *World Journal of Methodology* 11 (4): 116–29. https://doi.org/10.5662/wjm.v11.i4.116.

Nature Editorial. 2020. “Why Nature Needs to Cover Politics Now More than Ever.” *Nature* 586 (7828): 169–70. https://doi.org/10.1038/d41586-020-02797-1.

Nchasi, Goodluck, Innocent Kitandu Paul, Sospeter Berling Sospeter, Margareth Richard Mallya, Juvenali Ruaichi, and Johnson Malunga. 2022. “Measles Outbreak in Sub-Saharan Africa amidst COVID-19: A Rising Concern, Efforts, Challenges, and Future Recommendations.” *Annals of Medicine and Surgery* 81 (July): 104264. https://doi.org/10.1016/j.amsu.2022.104264.

Nelson, Fraser. 2022. “The Lockdown Files: Rishi Sunak on What We Weren’t Told.” *The Spectator*, August 25. https://www.spectator.co.uk/article/the-lockdown-files-rishi-sunak-on-what-we-werent-told/.

NHS, U.K. 2023. “Getting a COVID-19 Vaccine.” *Nhs.Uk*, March 20. https://www.nhs.uk/conditions/covid-19/covid-19-vaccination/getting-a-covid-19-vaccine/.

Nixon, Kristen, Sonia Jindal, Felix Parker, Nicholas G. Reich, Kimia Ghobadi, Elizabeth C. Lee, Shaun Truelove, and Lauren Gardner. 2022. “An Evaluation of Prospective COVID-19 Modelling Studies in the USA: From Data to Science Translation.” *The Lancet Digital Health* 4 (10): e738–47. https://doi.org/10.1016/S2589-7500(22)00148–0.

Nolen, Stephanie. 2022. “Trying to Solve a Covid Mystery: Africa’s Low Death Rates.” Health. *The New York Times*, March 23. https://www.nytimes.com/2022/03/23/health/covid-africa-deaths.html.

Northcott, Robert. 2022. “Pandemic Modeling, Good and Bad.” *Philosophy of Medicine* 3 (1): 1–20. https://doi.org/10.5195/POM.2022.79.

Nyashanu, Mathew, Prisca Simbanegavi, and Linda Gibson. 2020. “Exploring the Impact of COVID-19 Pandemic Lockdown on Informal Settlements in Tshwane Gauteng Province, South Africa.” *Global Public Health* 15 (10): 1443–53. https://doi.org/10.1080/17441692.2020.1805787.

Odenbaugh, Jay. 2011. “True Lies: Realism, Robustness, and Models.” *Philosophy of Science* 78 (5): 1177–88. https://doi.org/10.1086/662281.

OECD. 2022. *Education at a Glance 2022: OECD Indicators*. OECD Publishing. https://doi.org/10.1787/3197152b-en.

ONS, U.K. 2021. “Coronavirus and the social impacts on Great Britain: 25 June 2021.” *UK Office for National Statistics, June 25*. https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthandwellbeing/bulletins/coronavirusandthesocialimpactsongreatbritain/25june2021.

ONS, U.K. 2023. “Coronavirus (COVID-19) Infection Survey Technical Article: Cumulative Incidence of the Percentage of People Who Have Been Infected with COVID-19 by Variant and Age, England – Office for National Statistics.” *UK Office for National Statistics*, February 9. https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/articles/coronaviruscovid19infectionsurveytechnicalarticlecumulativeincidenceofthenumberofpeoplewhohavebeeninfectedwithcovid19byvariantandageengland/9february2023.

Our World in Data. 2024. “Cumulative Confirmed COVID-19 Cases and Deaths.” Our World in Data. https://ourworldindata.org/grapher/cumulative-deaths-and-cases-covid−19.

Oxfam. 2022. “Poor Countries Denied $5.7 Trillion in Aid Because of Rich Countries’ 50-Year Failure to Deliver on Aid Promises.” *Oxfam International*, May 25. https://www.oxfam.org/en/press-releases/poor-countries-denied-57-trillion-aid-because-rich-countries-50-year-failure-deliver.

Oxfam. 2024. “Why We Need a People’s Vaccine.” *Oxfam International*, February 13. https://www.oxfam.org/en/take-action/campaigns/covid-19-vaccine/about.

Pamuk, Zeynep. 2021. “Politics and Expertise: How to Use Science in a Democratic Society.” In *Politics and Expertise*. Princeton University Press. https://doi.org/10.1515/9780691218946.

Pamuk, Zeynep. 2022. “COVID-19 and the Paradox of Scientific Advice.” *Perspectives on Politics* 20 (2): 562–76. https://doi.org/10.1017/S1537592721001201.

Parker, Wendy S. 2006. “Understanding Pluralism in Climate Modeling.” *Foundations of Science* 11 (4): 349–68. https://doi.org/10.1007/s10699-005-3196-x.

Parker, Wendy S. 2010. “Whose Probabilities? Predicting Climate Change with Ensembles of Models.” *Philosophy of Science* 77 (5): 985–97. https://doi.org/10.1086/656815.

Parker, Wendy S. 2011. “When Climate Models Agree: The Significance of Robust Model Predictions | Philosophy of Science.” *Philosophy of Science* 78 (4): 579–600. https://doi.org/10.1086/661566.

Parker, Wendy S. 2020. “Model Evaluation: An Adequacy-for-Purpose View.” *Philosophy of Science* 87 (3): 457–77.

Parker, Wendy S., and Eric Winsberg. 2018. “Values and Evidence: How Models Make a Difference.” *European Journal for Philosophy of Science* 8 (1): 125–42. https://doi.org/10.1007/s13194-017-0180-6.

Patone, Martina, Xue W. Mei, Lahiru Handunnetthi, Sharon Dixon, Francesco Zaccardi, Manu Shankar-Hari, Peter Watkinson, et al. 2022. “Risks of Myocarditis, Pericarditis, and Cardiac Arrhythmias Associated with COVID-19 Vaccination or

SARS-CoV-2 Infection." *Nature Medicine* 28 (2): 410–22. https://doi.org/10.1038/s41591-021-01630-0.

Pearce, Warren. 2023. "What Does Covid-19 Mean for Expertise? The Case of Tomas Pueyo." December 15. https://www.sheffield.ac.uk/ihuman/blog/what-does-covid-19-mean-expertise-case-tomas-pueyo.

Pegden, Wesley. 2021. "Weighing Myocarditis Cases, ACIP Failed to Balance the Harms vs Benefits of 2nd Doses." *Medium*, June 24. https://medium.com/@wpegden/weighing-myocarditis-cases-acip-failed-to-balance-the-harms-vs-benefits-of-2nd-doses-d7d6b3df7cfb.

Pielke, Jr., Roger A. 2007. *The Honest Broker: Making Sense of Science in Policy and Politics*. Cambridge University Press. https://doi.org/10.1017/CBO9780511818110.

Piper, Kelsey. 2020. "This Coronavirus Model Keeps Being Wrong. Why Are We Still Listening to It?" *Vox*, May 2. https://www.vox.com/future-perfect/2020/5/2/21241261/coronavirus-modeling-us-deaths-ihme-pandemic.

Polver, Marco, Fabio Previdi, Mirko Mazzoleni, and Alberto Zucchi. 2021. "A SIAT3HE Model of the COVID-19 Pandemic in Bergamo, Italy." *IFAC-PapersOnLine, 11th IFAC Symposium on Biological and Medical Systems BMS 2021* 54 (15): 263–68. https://doi.org/10.1016/j.ifacol.2021.10.266.

Prasad, Vinay, and Alyson Haslam. 2024. "COVID-19 Vaccines: History of the Pandemic's Great Scientific Success and Flawed Policy Implementation." *Monash Bioethics Review* (March 9): 1–27. https://doi.org/10.1007/s40592-024-00189-z.

Pueyo, Tomas. 2020a. "Coronavirus: Why You Must Act Now." *Medium*, March 10. https://tomaspueyo.medium.com/coronavirus-act-today-or-people-will-die-f4d3d9cd99ca.

Pueyo, Tomas. 2020b. "Coronavirus: The Hammer and the Dance." *Medium*, March 19. https://tomaspueyo.medium.com/coronavirus-the-hammer-and-the-dance-be9337092b56.

Qualls, Noreen. 2017. "Community Mitigation Guidelines to Prevent Pandemic Influenza — United States, 2017." *MMWR Recommendations and Reports*. https://doi.org/10.15585/mmwr.rr6601a1.

Recht, Benjamin. 2021. "Revisiting the Bangladesh Mask RCT." *Arg Min Blog*. http://archives.argmin.net/2021/11/23/mask-rct-revisited/.

Remuzzi, Andrea, and Giuseppe Remuzzi. 2020. "COVID-19 and Italy: What Next?" *The Lancet* 395 (10231): 1225–28. https://doi.org/10.1016/S0140-6736(20)30627–9.

Reuters. 2021. "Israel Examining Heart Inflammation Cases in People Who Received Pfizer COVID Shot." *Reuters*, April 25. https://www.reuters.com/world/middle-east/israel-examining-heart-inflammation-cases-people-who-received-pfizer-covid-shot-2021-04-25/.

Reza, Faisal bin-, Vicente Lopez Chavarrias, Angus Nicoll, and Mary E. Chamberland. 2012. "The Use of Masks and Respirators to Prevent Transmission of Influenza: A Systematic Review of the Scientific Evidence." *Influenza and Other Respiratory Viruses* 6 (4): 257–67. https://doi.org/10.1111/j.1750-2659.2011.00307.x.

Robins, James M., Miguel A. Hernán, and Babette Brumback. 2000. "Marginal Structural Models and Causal Inference in Epidemiology." *Epidemiology* 11 (5): 550–60.

Rohban, Thomas. 2022. "BNT162b2 Vaccine Booster and Covid-19 Mortality." *New England Journal of Medicine* 386 (10): 1000–1. https://doi.org/10.1056/NEJMc2120044.

Rose, Geoffrey. 1992. *The Strategy of Preventive Medicine*. Oxford University Press.

The Royal Society. 2023. "COVID-19: Examining the Effectiveness of Non-Pharmaceutical Interventions." The Royal Society. https://royalsociety.org/-/media/policy/projects/impact-non-pharmaceutical-interventions-on-covid-19-transmission/the-royal-society-covid-19-examining-the-effectiveness-of-non-pharmaceutical-interventions-report.pdf.

SACMC. 2020. "Estimating Cases for COVID-19 in South Africa: Short Term Projections: June 2020." South African Covid-19 Modelling Consortium. https://www.nicd.

ac.za/wp-content/uploads/2020/06/SACovidModellingReport_ShortTermProjections_12062020_Final2.pdf.
SACMC. 2021. "SACMC Fourth Wave Report." South Africa Covid-19 Modelling Consortium. https://www.nicd.ac.za/wp-content/uploads/2021/11/SACMC-Fourth-wave-report-17112021-final.pdf.
SAGE. 2021a. "SPI-M-O: Summary of Further Modelling of Easing Restrictions – Roadmap Step 4 on 19th July 2021." Scientific Advisory Group for Emergencies (UK). https://www.gov.uk/government/publications/spi-m-o-summary-of-further-modelling-of-easing-restrictions-roadmap-step-4-on-19-july-2021–7-july-2021/spi-m-o-summary-of-further-modelling-of-easing-restrictions-roadmap-step-4-on-19-july-2021–7-july–2021.
SAGE. 2021b. "SPI-M-O: Consensus Statement on COVID-19, 15 December 2021." *GOV.UK*, December 15. https://www.gov.uk/government/publications/spi-m-o-consensus-statement-on-covid-19–15-december-2021/spi-m-o-consensus-statement-on-covid-19–15-december–2021.
Sandlund, Johanna, Ram Duriseti, Shamez N. Ladhani, Kelly Stuart, Jeanne Noble, and Tracy Beth Høeg. 2023. "Child Mask Mandates for COVID-19: A Systematic Review." *Archives of Disease in Childhood* (December 2): ahead of print. https://doi.org/10.1136/archdischild-2023-326215.
Sayers, Freddie. 2020. "Neil Ferguson Interview: China Changed What Was Possible." *UnHerd*, December 26. https://unherd.com/thepost/neil-ferguson-interview-china-changed-what-was-possible/.
Schoch, Marta, and Christoph Lakner. 2020. "The Number of Poor People Continues to Rise in Sub-Saharan Africa, despite a Slow Decline in the Poverty Rate." December 16. https://blogs.worldbank.org/opendata/number-poor-people-continues-rise-sub-saharan-africa-despite-slow-decline-poverty-rate.
Schwalb, Alvaro, and Carlos Seas. 2021. "The COVID-19 Pandemic in Peru: What Went Wrong?" *The American Journal of Tropical Medicine and Hygiene* 104 (4): 1176–78. https://doi.org/10.4269/ajtmh.20-1323.
Science Media Centre. 2020. "Expert Reaction to Paper Using an RCT to Assess Mask Use as a Public Health Measure to Help Control SARS-CoV-2 Spread (DANMASK-19)." *Science Media Centre*, November 19. https://www.sciencemediacentre.org/expert-reaction-to-paper-using-an-rct-to-assess-mask-use-as-a-public-health-measure-to-help-control-sars-cov-2-spread-danmask-19/.
Sebhatu, Abiel, Karl Wennberg, Stefan Arora-Jonsson, and Staffan I. Lindberg. 2020. "Explaining the Homogeneous Diffusion of COVID-19 Nonpharmaceutical Interventions across Heterogeneous Countries." *Proceedings of the National Academy of Sciences* 117 (35): 21201–8. https://doi.org/10.1073/pnas.2010625117.
Shan, Yafeng, and Jon Williamson. 2023. *Evidential Pluralism in the Social Sciences*. Taylor & Francis.
Sharrow, David, Lucia Hug, Danzhen You, Leontine Alkema, Robert Black, Simon Cousens, Trevor Croft, et al. 2022. "Global, Regional, and National Trends in under-5 Mortality between 1990 and 2019 with Scenario-Based Projections until 2030: A Systematic Analysis by the UN Inter-Agency Group for Child Mortality Estimation." *The Lancet Global Health* 10 (2): e195–206. https://doi.org/10.1016/S2214-109X(21)00515–5.
Silberzahn, R., E.L. Uhlmann, D.P. Martin, Pasquale Anselmi, Frederik Aust, Eli Awtrey, Štěpán Bahník, et al. 2018. "Many Analysts, One Data Set: Making Transparent How Variations in Analytic Choices Affect Results." *Advances in Methods and Practices in Psychological Science* 1 (3): 337–56. https://doi.org/10.1177/2515245917747646.
Silver, Nate. 2012. *The Signal and the Noise: The Art and Science of Prediction*. Penguin UK.

Smart, Benjamin T.H., and Alex Broadbent. 2020. "South Africa's COVID-19 Lockdown: Cigarettes and Outdoor Exercise Could Ease the Tension." *The Conversation*, March 30. http://theconversation.com/south-africas-covid-19-lockdown-cigarettes-and-outdoor-exercise-could-ease-the-tension−134931.

Smith, Louise, Henry Potts, Richard Amlôt, Nicola Fear, Susan Michie, and James Rubin. 2021. "Adherence to the test, trace, and isolate system in the UK: results from 37 nationally representative surveys." *BMJ* 372 (December): n608. https://doi.org/10.1136/bmj.n608.

Soltesz, Kristian, Fredrik Gustafsson, Toomas Timpka, Joakim Jaldén, Carl Jidling, Albin Heimerson, Thomas B. Schön, et al. 2020. "The Effect of Interventions on COVID-19." *Nature* 588 (7839): 7839. https://doi.org/10.1038/s41586-020-3025-y.

Straus, Sharon E., Paul Glasziou, W. Scott Richardson, and R. Brian Haynes. 2018. *Evidence-Based Medicine*. 5th ed. Elsevier. https://shop.elsevier.com/books/evidence-based-medicine/straus/978-0-7020-6296-4.

Streicher, Pieter. 2021. "Omicron Analysis." Tweet. *Twitter*, December 19. https://twitter.com/pieterstreicher/status/1472447879088386059.

Streicher, Pieter, and Alex Broadbent. 2023. Pandemic response strategies and threshold phenomena. *Global Epidemiology* 5 (December): 100105.

Streicher, Pieter, Alex Broadbent, and Joel Hellewell. 2025. "The Need for Methodological Pluralism in Epidemiological Modelling." *Global Epidemiology* 9 (June): 100177. https://doi.org/10.1016/j.gloepi.2024.100177.

Suárez, Mauricio. 2024. *Inference and Representation: A Study in Modeling Science*. University of Chicago Press.

Taylor, Chloe. 2021. "Nordic Countries Are Restricting the Use of Moderna's Covid Vaccine. Here's Why." *CNBC*, October 8. https://www.cnbc.com/2021/10/08/nordic-countries-are-restricting-the-use-of-modernas-covid-vaccine.html.

Team, IHME COVID-19 Health Service Utilization Forecasting, and Christopher J.L. Murray. 2020. "Forecasting COVID-19 Impact on Hospital Bed-Days, ICU-Days, Ventilator-Days and Deaths by US State in the next 4 Months." *Preprint*, *medRxiv*, March 30. https://doi.org/10.1101/2020.03.27.20043752.

Teller, Paul. 2001. "Twilight of the Perfect Model." *Erkenntnis (1975-)* 55 (3): 393–415.

Tetlock, Philip. 2005. *Expert Political Judgement – How Good Is It? How Can We Know?* Princeton University Press.

Thomas, Stephen J., Edson D. Moreira, Nicholas Kitchin, Judith Absalon, Alejandra Gurtman, Stephen Lockhart, John L. Perez, et al. 2021. "Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine through 6 Months." *The New England Journal of Medicine* 385 (19): 1761–73. https://doi.org/10.1056/NEJMoa2110345.

UK Covid-19 Inquiry. 2023a. "Transcript of Module 2 Public Hearing on 12 October 2023 UK Covid-19 Inquiry Archives." UK Covid-19 Inquiry, October 12. https://covid19.public-inquiry.uk/documents/transcript-of-module-2-public-hearing-on-12-october-2023/.

UK Covid-19 Inquiry. 2023b. "Transcript of Module 2 Public Hearing on 16 October 2023." UK Covid-19 Inquiry, October 16. https://covid19.public-inquiry.uk/documents/transcript-of-module-2-public-hearing-on-16-october-2023/.

UK Covid-19 Inquiry. 2023c. "Transcript of Module 2 Public Hearing on 17 October 2023 UK Covid-19 Inquiry Archives." UK Covid-19 Inquiry, October 17. https://covid19.public-inquiry.uk/documents/transcript-of-module-2-public-hearing-on-17-october-2023/.

UK Govt Coronavirus Data. 2024. "Cases in England | Coronavirus in the UK." https://coronavirus.data.gov.uk/details/cases?areaType=nation&areaName=England.

UN. 2021. "Pandemic Year Marked by Spike in World Hunger." United Nations. https://www.who.int/news/item/12-07-2021-un-report-pandemic-year-marked-by-spike-in-world-hunger.
UN Department of Economic and Social Affairs. 2023. "Goal 11." Sustainable Development Goals. https://sdgs.un.org/goals/goal11.
UN Habitat. 2007. "What Are Slums and Why Do They Exist?" UN Habitat. https://www.preventionweb.net/files/1700_462551419GC202120What20are20slums.pdf.
UN Women. 2020. "The Shadow Pandemic: Violence against Women during COVID-19." UN Women – Headquarters. https://www.unwomen.org/en/news/in-focus/in-focus-gender-equality-in-covid-19-response/violence-against-women-during-covid−19.
UNESCO. 2013. "Education Transforms Lives – UNESCO Digital Library." UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000223115.
UNESCO. 2020. "Keeping Girls in the Picture." UNESCO. https://www.unesco.org/en/covid-19/education-response/keeping-girls-picture.
UNICEF. 2020. "At Least 80 Million Children under One at Risk of Diseases Such as Diphtheria, Measles and Polio as COVID-19 Disrupts Routine Vaccination Efforts, Warn Gavi, WHO and UNICEF." May 26. https://www.unicef.org/romania/press-releases/least-80-million-children-under-one-risk-diseases-such-diphtheria-measles-and-polio.
UNSDG. 2020. "UNSDG | Policy Brief: The Impact of COVID-19 on Women." United Nations Sustainable Development Group. https://unsdg.un.org/resources/policy-brief-impact-covid-19-women.
Vandenbroucke, Jan P., Alex Broadbent, and Neil Pearce. 2016. "Causality and Causal Inference in Epidemiology: The Need for a Pluralistic Approach." *International Journal of Epidemiology* 45 (6): 1776–86. https://doi.org/10.1093/ije/dyv341.
Verity, Robert, Lucy C. Okell, Ilaria Dorigatti, Peter Winskill, Charles Whittaker, Natsuko Imai, Gina Cuomo-Dannenburg, et al. 2020. "Estimates of the Severity of COVID-19 Disease." *Preprint, medRxiv*, March 13. https://doi.org/10.1101/2020.03.09.20033357.
Viner, Russell M., Simon J. Russell, Helen Croker, Jessica Packer, Joseph Ward, Claire Stansfield, Oliver Mytton, Chris Bonell, and Robert Booy. 2020. "School Closure and Management Practices during Coronavirus Outbreaks Including COVID-19: A Rapid Systematic Review." *The Lancet Child and Adolescent Health* 4 (5): 397–404. https://doi.org/10.1016/S2352-4642(20)30095-X.
Wadud, Zia, Sheikh Mokhlesur Rahman, and Annesha Enam. 2022. "Face Mask Mandates and Risk Compensation: An Analysis of Mobility Data during the COVID-19 Pandemic in Bangladesh." *BMJ Global Health* 7 (1). https://doi.org/10.1136/bmjgh-2021-006803.
Walker, Patrick Gt, Charles Whittaker, Oliver Watson, K. Ainslie, and S. Bhatia. 2020. "Report 12: The Global Impact of COVID-19 and Strategies for Mitigation and Suppression." Imperial College London. https://doi.org/10.25561/77735.
Watanabe, Tsutomu, and Tomoyoshi Yabu. 2021. "Japan's Voluntary Lockdown: Further Evidence Based on Age-Specific Mobile Location Data." *Japanese Economic Review (Oxford, England)* 72 (3): 333–70. https://doi.org/10.1007/s42973-021-00077-9.
Watson, Oliver J., Gregory Barnsley, Jaspreet Toor, Alexandra B. Hogan, Peter Winskill, and Azra C. Ghani. 2022. "Global Impact of the First Year of COVID-19 Vaccination: A Mathematical Modelling Study." *The Lancet Infectious Diseases* 22 (9): 1293–302. https://doi.org/10.1016/S1473-3099(22)00320–6.
WEF. 2024. "Global Risks Report 2024." Insight Report No. 19. Global Risks Reports. World Economic Forum. https://www.weforum.org/publications/global-risks-report-2024/.
Weisberg, Michael. 2013. *Simulation and Similarity: Using Models to Understand the World*. OUP USA.

Wenstrup, Brad. 2024. "Wenstrup Issues Key Takeaways from Interview with Dr. Francis Collins." United States House Committee on Oversight and Accountability, January 13. https://oversight.house.gov/release/wenstrup-issues-key-takeaways-from-interview-with-dr-francis-collins/.

Whipple, Tom. 2022. "Sage Modellers Admit They Got Omicron Prediction Wrong." *The Times*, March 3. https://www.thetimes.com/uk/science/article/sage-modellers-admit-they-got-omicron-prediction-wrong-ck8mhl5g7.

White, Lucie, Philippe van Basshuysen, and Mathias Frisch. 2022. "When Is Lockdown Justified?" *Philosophy of Medicine* 3 (1): 1–22. https://doi.org/10.5195/POM.2022.85.

Whitty, Christopher J.M. 2015. "What Makes an Academic Paper Useful for Health Policy?" *BMC Medicine* 13 (1): 301. https://doi.org/10.1186/s12916-015-0544-8.

WHO. 2008. "Closing the Gap in a Generation: Health Equity through Action on the Social Determinants of Health – Final Report of the Commission on Social Determinants of Health." WHO Commission on Social Determinants of Health. https://www.who.int/publications/i/item/WHO-IER-CSDH−08.1.

WHO. 2019. "Non-Pharmaceutical Public Health Measures for Mitigating the Risk and Impact of Epidemic and Pandemic Influenza." World Health Organization. https://iris.who.int/handle/10665/329438.

WHO. 2020a. "Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19)." World Health Organization. https://www.who.int/publications-detail-redirect/report-of-the-who-china-joint-mission-on-coronavirus-disease-2019-(covid-19).

WHO. 2020b. "The Six Steps." World Health Organization, April. https://cdn.who.int/media/docs/default-source/searo/thailand/brief-six-criteria-that-each-country-should-meet-before-lifting-restrictions.pdf.

WHO. 2023. "Global Excess Deaths Associated with COVID-19 (Modelled Estimates)." World Health Organisation. https://www.who.int/data/sets/global-excess-deaths-associated-with-covid-19-modelled-estimates.

WHO. 2024a. "About WHO: Collaborating Centres." World Health Organisation. https://www.who.int/about/collaboration/collaborating-centres.

WHO. 2024b. "Detail of Infectious Disease Modelling WHO Collaborating Centre." World Health Organisation. https://apps.who.int/whocc/Detail.aspx?FJdjEEdrcMNfOU4d+dseSg==.

WHO. 2024c. "WHOCC – WHO Collaborating Centres." World Health Organisation. https://apps.who.int/whocc/List.aspx?QlbF+x5BZXHl4ya3zHVRnIErSqPVgFzujE5PrGEi1kI=.

WHO. 2024d. "Fact Sheets – Malnutrition." World Health Organisation, March 1. https://www.who.int/news-room/fact-sheets/detail/malnutrition.

Wighton, Kate. 2020. "Lockdown and School Closures in Europe May Have Prevented 3.1m Deaths | Imperial News | Imperial College London." *Imperial News*, June 8. https://www.imperial.ac.uk/news/198074/lockdown-school-closures-europe-have-prevented/.

Winsberg, Eric. 2010. *Science in the Age of Computer Simulation*. University of Chicago Press. https://press.uchicago.edu/ucp/books/book/chicago/S/bo9003670.html.

Winsberg, Eric, Jason Brennan, and Chris W. Surprenant. 2020. "How Government Leaders Violated Their Epistemic Duties During the SARS-CoV-2 Crisis." *Kennedy Institute of Ethics Journal* 30 (3): 215–42.

Winsberg, Eric, and Stephanie Harvard. 2022. "Purposes and Duties in Scientific Modelling." Essay. *J Epidemiol Community Health* 76 (5): 512–17. https://doi.org/10.1136/jech-2021-217666.

Winsberg, Eric, and Stephanie Harvard. 2024. *Scientific Models and Decision Making*. Elements in the Philosophy of Science. Cambridge University Press. Cambridge Core. https://doi.org/10.1017/9781009029346.

Wood, Simon N., and Ernst C. Wit. 2021. “Was R < 1 before the English Lockdowns? On Modelling Mechanistic Detail, Causality and Inference about Covid-19.” *PLoS ONE* 16 (9): e0257455. https://doi.org/10.1371/journal.pone.0257455.

Workman, Herbert B. 1926. *John Wyclif: A Study of the English Medieval Church*. Oxford University Press.

World Bank. 2018. “World Development Report 2018: Learning to Realize Education’s Promise.” World Bank. https://doi.org/10.1596/978-1-4648-1096-1.

The World Bank, UNESCO, and UNICEF. 2021. “The State of the Global Education Crisis: A Path to Recovery.” No. 166631. International Bank for Reconstruction and Development/The World Bank, UNESCO and UNICEF. https://documents.worldbank.org/en/publication/documents-reports/documentdetail/416991638768297704/The-State-of-the-Global-Education-Crisis-A-Path-to-Recovery.

Worldometer. 2024. “Population of Africa (2024).” https://www.worldometers.info/world-population/africa-population/.

Yorke, James A., Herbert W. Hethcote, and Annett Nold. 1978. “Dynamics and Control of the Transmission of Gonorrhea.” *Sexually Transmitted Diseases* 5 (2): 51–56. JSTOR.

INDEX