



OPEN ACCESS

Adapt or die: how the pandemic made the shift from EBM to EBM+ more urgent

Trisha Greenhalgh ¹, David Fisman,² Danielle J Cane,³ Matthew Oliver ⁴, Chandini Raina Macintyre⁵

10.1136/bmjebm-2022-111952

¹Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, UK

²Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada

³Coalition for Healthcare Acquired Infection Reduction, Cambridge, Ontario, Canada

⁴Association of Professional Engineers and Geoscientists, Edmonton, Alberta, Canada

⁵The Biosecurity Program, Kirby Institute, University of New South Wales, Sydney, New South Wales, Australia

Correspondence to:

Professor Trisha Greenhalgh, Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford OX2 6GG, UK; trish.greenhalgh@phc.ox.ac.uk



Check for updates

© Author(s) (or their employer(s)) 2022. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

To cite: Greenhalgh T, Fisman D, Cane DJ, *et al.* *BMJ Evidence-Based Medicine* Epub ahead of print: [please include Day Month Year]. doi:10.1136/bmjebm-2022-111952

Abstract

Evidence-based medicine (EBM's) traditional methods, especially randomised controlled trials (RCTs) and meta-analyses, along with risk-of-bias tools and checklists, have contributed significantly to the science of COVID-19. But these methods and tools were designed primarily to answer simple, focused questions in a stable context where yesterday's research can be mapped more or less unproblematically onto today's clinical and policy questions. They have significant limitations when extended to complex questions about a novel pathogen causing chaos across multiple sectors in a fast-changing global context. Non-pharmaceutical interventions which combine material artefacts, human behaviour, organisational directives, occupational health and safety, and the built environment are a case in point: EBM's experimental, intervention-focused, checklist-driven, effect-size-oriented and deductive approach has sometimes confused rather than informed debate. While RCTs are important, exclusion of other study designs and evidence sources has been particularly problematic in a context where rapid decision making is needed in order to save lives and protect health. It is time to bring in a wider range of evidence and a more pluralist approach to defining what counts as 'high-quality' evidence. We introduce some conceptual tools and quality frameworks from various fields involving what is known as mechanistic research, including complexity science, engineering and the social sciences. We propose that the tools and frameworks of mechanistic evidence, sometimes known as 'EBM+' when combined with traditional EBM, might be used to develop and evaluate the interdisciplinary evidence base needed to take us out of this protracted pandemic. Further articles in this series will apply pluralistic methods to specific research questions.

Introduction

Two and a half years into the COVID-19 pandemic, it is time to take stock. What began as—we assumed—an acute respiratory illness rapidly sweeping the world has become a prolonged global crisis with medical, social, economic and political dimensions. The SARS-CoV-2 virus has proved tenacious and shifting. It causes a complex multi-system disease which disproportionately impacts minority ethnic groups and the sick, poor, old and disadvantaged. It produces significant mortality

and—in some—prolonged sequelae. Effective and safe vaccines were produced rapidly, but uptake has been patchy and highly transmissible variants continue to spread and mutate. Coordinated disinformation campaigns have weakened the public health response.

Despite a quarter of a million scientific papers on COVID-19, some basic issues remain contested. How exactly does the virus spread? How effective are non-pharmaceutical interventions—masks, distancing, closure of buildings, remote working and learning, lockdowns—in reducing transmission, and what are their trade-offs? How can we make schools, hospitals and other public buildings safe? How can we protect workers and the public without closing down the economy? How can we reduce the shocking inequalities that have characterised this pandemic?

This paper explains why we need to go beyond lip-service to evidential pluralism to fully answer those questions. While acknowledging the primacy of randomised controlled trials (RCTs) and meta-analyses of RCTs for estimating the efficacy of drugs and vaccines, we highlight and extend the arguments made by some scholars within the evidence-based medicine (EBM) movement for 'EBM+',^{1,2} defined as an approach which systematically considers mechanistic evidence (studies which aim to explain which factors and interactions are responsible for a phenomenon³) on a par with probabilistic clinical and epidemiological studies. Our central argument is that for some aspects of the pandemic, especially those characterised by a combination of complexity (multiple variables interacting dynamically with a high degree of uncertainty), urgency (decisions needed in days not years) and threat (the consequences of not acting could be catastrophic), mechanistic evidence has been mission-critical and RCTs difficult or impossible. Thousands of lives were likely lost as a result of what was incorrectly claimed to be an "evidence-based" approach—dismissing or downgrading mechanistic evidence, overvaluing findings from poorly designed or irrelevant RCTs, and advocating for inaction where RCT evidence was lacking. The pandemic is an epistemic opportunity for the EBM movement to come to better understand, debate and embrace EBM+.

Evidence hierarchies: not fit for all purposes

Every student of EBM is taught that the hierarchy of evidence—with meta-analyses of RCTs at the top and so-called 'less trustworthy' evidence

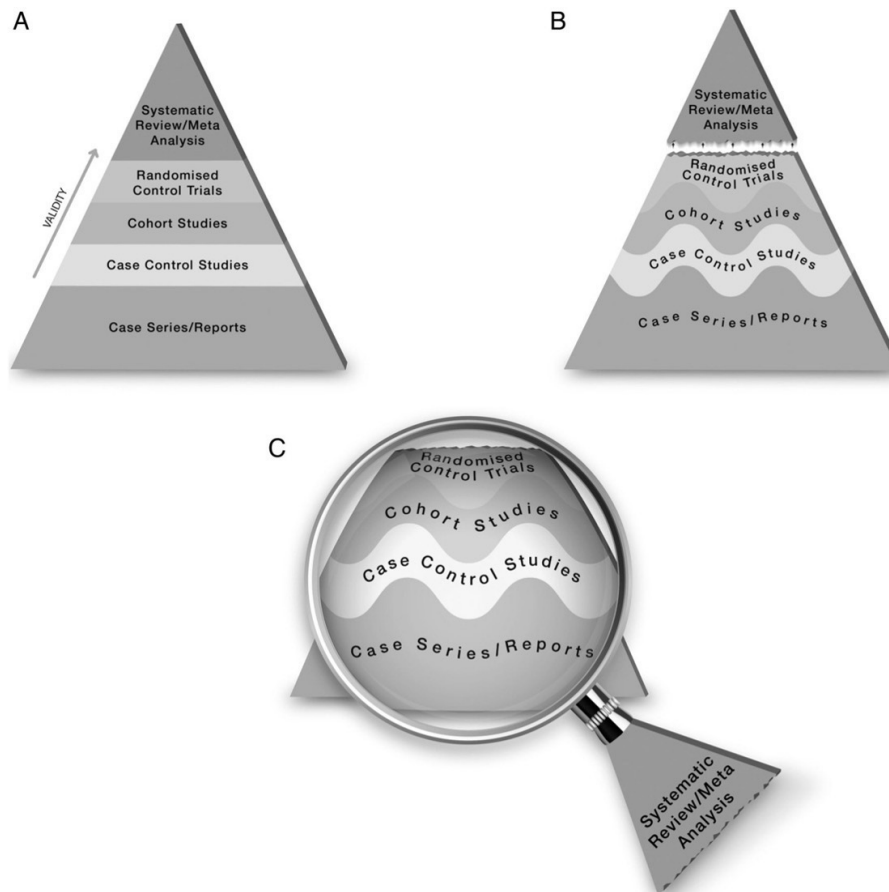


Figure 1 One version of EBM's hierarchy of evidence, showing evolution over time to encourage a more flexible approach. (Reproduced under CC-BY-4.0 Licence from Murad *et al*⁴). EBM, evidence-based medicine.

(case studies, mechanistic evidence, expert opinion) at the bottom (figure 1A)⁴—is a useful heuristic but does not apply in all circumstances. But guidance to apply the hierarchy flexibly to take account of—for example—the nature of the research question, the degree of complexity involved, whether RCTs of the necessary size and scope are actually feasible, and whether policy decisions can or should be put on hold pending definitive results—has been honoured more in the breach than the observance. From Guyatt *et al*'s⁵ paper announcing EBM as a 'new paradigm' in medicine to their 2008 Grading of Recommendations, Assessment, Development and Evaluation statement (which offers eight domains of strengths and limitations of the overall evidence base, including observational studies as well as RCTs)⁶ and the Oxford Centre for EBM 2011 Levels of Evidence Document,⁷ there remains within EBM orthodoxy an explicitly *hierarchical* relationship between probabilistic clinical and epidemiological evidence (invariably placed at the top of the hierarchy) and the kinds of evidence which help establish causal mechanisms (invariably placed at or near the bottom).

The cornerstone of EBM as taught to students—ask focused questions in population-intervention (or exposure)-comparison-outcome (PICO) format; search for RCTs and, if these are unavailable, work down the hierarchy of evidence; appraise studies for risk of bias; and combine them using meta-analysis—was developed for a stable context in which predictions based on yesterday's research map unproblematically onto today's policy questions (eg, what actions should we take to prevent and manage coronavirus?). Despite their theoretical strengths for some kinds of

question, and despite heroic achievements in the pandemic with fast-track platform trials such as RECOVERY,⁸ PRINCIPLE⁹ and PANORAMIC,¹⁰ RCTs are not a panacea (table 1).¹¹ Modifications of the hierarchy of evidence (eg, by separating the layers with dotted or wavy lines, as in figure 1B, and using systematic review as a lens through which primary evidence is interpreted, as in figure 1C⁴) fall short of the epistemic shift that is now needed.

Strengthening causal claims through mechanistic evidence

Mechanistic evidence includes a wide range of designs, including in vitro experiments, biomedical imaging, autopsy, established theory, animal experiments, aerosol science, engineering research and simulations.³ Like all evidence, mechanistic studies can be of variable quality and definitiveness. Table 2 shows some preliminary criteria for grading mechanistic evidence, based partly on previous studies,¹²⁻¹⁴ which would benefit from further elaboration.

No one kind of evidence stands alone. Rather, as encapsulated in the Bradford Hill indicators¹⁵⁻¹⁷ (table 3), evidence of mechanisms must be combined with probabilistic evidence from clinical trials (and, where appropriate, evidence from non-randomised comparative and observational studies) to make a strong case for causality.¹³ This is because real-world circumstances often differ from the ones in which a probabilistic estimate (such as an effect size in an RCT) was demonstrated; evidence supporting a plausible mechanism affirms (or not) that the causal relationship is likely to be stable across settings.¹³ The presence of high-quality and consistent mechanistic evidence greatly increases the external

Table 1 Strengths and limitations of RCTs and meta-analyses for pandemic research

'Quality' features of RCTs and meta-analyses of RCTs	Reason why these designs are sometimes problematic in a pandemic
<ul style="list-style-type: none"> ▶ Randomisation, if successful, will ensure that comparisons between exposed and unexposed groups are fair—that is, that groups are effectively identical in all respects except for their exposure to the drug, vaccine or public health intervention in question. ▶ Clinical equipoise and informed consent helps ensure that a trial is ethical. ▶ Factorial, pick-the-winner, ring vaccination and other advanced RCT designs could potentially add helpful information about complexity and speed the generation of knowledge in a crisis. ▶ Meta-analysis allows objective and systematic assessment of individual trials and appropriate weighting by risk of bias. ▶ 'Living' systematic reviews can provide timely summaries of RCT evidence, as occurred with drugs and vaccines. 	<ul style="list-style-type: none"> ▶ Because communicable diseases (by definition) spread through communities, it may be impossible to ensure that cross-contamination of arms does not occur. ▶ Even rapid RCTs take time to plan and execute; in most cases even a rapidly conducted RCT will generate less timely information than real-time analysis of observational data already in hand. ▶ It may be impossible to power trials adequately to evaluate the full range of interactions of interest. ▶ It may be unethical to avoidably expose study participants to a virulent pathogen ▶ RCTs are not the preferred study design for questions of aetiology (eg, 'how does the SARS-CoV-2 virus spread?') ▶ Meta-analysis involves subjective judgments by the reviewer about how rigorously an effect was measured and how well context and confounders were taken into account in primary studies. ▶ If the individual RCTs in a meta-analysis include unfair comparisons, or if their findings are incorrectly combined, flawed summaries, false reassurance and inappropriate policy could result.
RCTs, randomised controlled trials.	

validity of an RCT finding; absence of such evidence should make us question RCT findings.³ Public health interventions involve two kinds of mechanisms: the upstream causes of disease (for COVID-19 'mechanisms involving family structures and interaction patterns, occupational behaviour, urban density, housing occupation and overcrowding, workplace and retail environment structures and organisation, as well as local social, economic and cultural variation'—Aronson JK, page 688² and the 'causes' through which an interventions might act (eg, attitudes, beliefs, capability, personal resources).²

Mechanistic evidence is inherently explanatory. It helps address the question 'how might this effect be produced?'. In the early months of the pandemic, the grammar of our research was prematurely fixed and PICO-constrained (eg, 'what is the effect size of masks in preventing respiratory infections?'). This narrowed our focus onto probabilistic studies of interventions ('masks-on'), comparisons ('masks-off') and predefined outcomes (infections in the wearer) expressed in statistical terms (see masking example in [box 1](#)¹⁸⁻²⁸). We suppressed our scientific imagination. We failed sufficiently to wonder at the novelty of the disease and the significance of its unique patterns of spread such as super-spreading events, overdispersion, transmission in the absence of symptoms, and indoor predominance—all of which should have raised mechanistic hypotheses about a predominantly airborne mechanism of transmission.²⁹

When the pandemic began, we immediately sought 'robust' evidence from new RCTs.³⁰ But overapplication of this heuristic led to narrowly focused reviews,³¹ excluded experts from other

disciplines, and closed our minds to the mechanistic explanations we could have built from well-conducted laboratory studies,³² animal studies,³³ modelling studies,³⁴ engineering studies³⁵ or careful analysis of real-world events³⁶—not to mention pre-pandemic studies on the transmission of comparable respiratory viruses.³⁷

Dynamic mechanisms: complex interventions in complex systems

A complex intervention (eg, facial protection with masks or respirators) consists of multiple components acting independently and inter-dependently at multiple levels (eg, material artefact, person, group, organisation, system), such that it is difficult to identify the 'active ingredient'.³⁸ Facial protection can be home-made or produced to formal technical standards; humans may agree—or not—to mask; masking may be socially expected, organisationally required or legally mandated; the threshold case rate for introducing or reintroducing masking may vary; global supply may be unreliable; masking is often paired with other interventions, introducing confounders.

A complex system—such as the one for delivering public health (which includes government, health service providers and the public)—is characterised by a number of features:

- ▶ Emergence (the system evolves over time, in ways that cannot be fully anticipated).
- ▶ Adaptation (the system adjusts to accommodate change).
- ▶ Feedback (either positive, where change in one factor reinforces another factor, or negative, where change in one factor diminishes another).

Table 2 A suggested hierarchy of evidence for mechanistic evidence (drawing partly on previous publications)^{12,14}

Level 1 (strongest)	Necessary and sufficient conditions for causality (eg, multiple features of the causal chain) supported by multiple independent studies, confirmed by multiple independent research groups using accepted best research methods. No high-quality disconfirming studies found.
Level 2	Indicators of causality (eg, more than one feature of the causal chain) supported, especially if by more than one method and confirmed by independent studies. No high-quality disconfirming studies found.
Level 3	Suggestion of causality (eg, one feature of the causal chain) supported by multiple independent studies. No high-quality disconfirming studies found.
Level 4	Sparse evidence supporting feature(s) of the causal chain. Disconfirming studies found but these are not definitive.
Level 5 (weakest)	No supporting studies. High-quality disconfirming studies found.

Table 3 Bradford-Hill indicators of causality, showing the importance of both probabilistic evidence (eg, from RCTs) and mechanistic evidence (eg, from laboratory studies)

Indicator	Explanation	Type	Preferred types of evidence
Strength of association	A strong association is more likely to be causal than a weak one	Both	RCTs or unbiased observational studies (eg, longitudinal cohort) which allow calculation of a measure of association such as relative risk or OR, and adjustment for confounding.
Consistency	Multiple observations made by different observers with different instruments mean the association is more likely to be causal	Both	Systematic review. Exploration of heterogeneity via tools such as meta-regression may help identify and quantify inconsistency in effects across studies.
Specificity	If an outcome is best predicted by one primary factor, the causal claim is more credible	Neither	A problematic indicator, as it is common for a single exposure to be causally associated with multiple outcomes (eg, tobacco smoking and heart disease, stroke, cancer, etc).
Temporality	A cause must precede an effect	Both	Any longitudinal design for example, clinical trial, non-cross-sectional observational study.
Biological gradient	There should be a direct 'dose-response' relationship between the independent variable (eg, a risk factor) and the dependent one (eg, people's status on the disease variable)	Both	Basic science (eg, toxicology-based risk analysis), varying-dose RCTs, unbiased observational studies with adjustment for confounding.
Plausibility	An association is more likely to be causal if there is a rational and theoretical basis for it	Mechanistic	Basic science
Coherence	An association is more likely to be causal if it coheres with other knowledge (ie, does not conflict with what is known about the variables under study and there are no plausible competing theories or rival hypotheses)	Both	Systematic review; integration of both mechanistic AND probabilistic evidence is key here.
Experimental manipulation	Any related research that is based on experiments will make a causal inference more plausible	Both	Basic science or RCT (which may be considered 'epidemiological experiments', as exposure is defined by the investigator).
Analogy	Sometimes a commonly accepted phenomenon in one area can be applied to another area	Both	Requires a broad understanding of all relevant fields; potentially subject to logical fallacy.

Adapted from Hill and Bradford Hill,¹⁵ incorporating important subsequent insights from Russo and Williamson¹⁶ and Fedak *et al.*¹⁷
RCT, randomised controlled trial.

- Self-organisation (eg, a public health department adapts a national programme to suit local constraints and priorities).³⁹

Particular outcomes emerge as a result of multiple interactions in the system.

Whereas studies of drug and vaccine efficacy can be interpreted within the traditional biomedical paradigm (table 4, column 2),⁴⁰ findings from such trials may not map to complex, real-world settings for multiple social and behavioural reasons. Studies how complex interventions generate outcomes in complex systems require a new paradigm (table 4, column 3) with designs that can capture dynamic change, accommodate non-linearity and embrace uncertainty.⁴¹

One approach to researching these generative mechanisms is the in-depth case study, a mixed-method design in which data collection, interpretation and synthesis are oriented to producing a rich description of a phenomenon in context.⁴² Hypotheses are explored and tested by constructing a plausible story (narrative synthesis) in which all data—qualitative and quantitative—are accounted for. Local public health interventions, even when part of national policy, are typically path-dependent (shaped and constrained by what has happened here previously) and emergent (they grow organically over time); they tend to be implemented and improved through testing and iteration in the real world.⁴³ As Ogilvie *et al.*⁴³ have pointed out, this pragmatic approach is not 'weaker' than an RCT but complementary to it (figure 2). Observations from real-world case studies inform hypothesis-driven questions about the impact of particular interventions under controlled conditions, which can be tested in RCTs. Conversely, RCTs may produce important findings which inform development and real-world testing of an intervention. These authors contrast the 'brick wall' of a meta-analysis (in which every primary study contributes a similar 'brick') with the 'dry stone wall' of a mixed-method

narrative review⁴⁴ (in which different kinds of primary evidence are combined to illuminate a problem from multiple angles).⁴⁵

Generative mechanisms in complex systems can also be studied quantitatively using modelling. Models (simplified versions of reality) can be used for a variety of purposes—to describe epidemic growth, estimate the likely impacts of interventions, simulate airflow inside buildings, and synthesise data from multiple sources in a manner analogous to systematic review and meta-analysis.⁴⁶ Model projections can be useful for presenting plausible counterfactuals, and for quantifying the impact of uncertainty. Policy-makers' use of modelling studies permit considering multiple 'what if' scenarios and sensitivity assessment as policy options develop. For communicable disease outbreaks, models can capture the non-independence of events that are the defining attribute of these diseases, and allow quantification of indirect effects on individuals who do not receive a given intervention⁴⁶—for example creation of herd effects via immunisation.

As with all research designs, real-world case studies^{42 47} and modelling⁴⁶ can be done well or badly, and the findings of such studies can be more or less relevant to a policy decision that is being contemplated in a different context or setting.

Mechanisms in the physical world: standards as evidence

Engineers use knowledge of the natural world to craft designs which benefit and improve society. That design process takes account of physical and technical properties and capabilities along with social needs and assessment of impact. Central to the science of engineering is a high reliance on standards, which are continuously developed to improve on the state of the art.⁴⁸ Indeed, quality in engineering evidence is defined in terms of relevant accepted standards and codes (eg, on indoor air quality controls for healthcare facilities).⁴⁹

Box 1 Public masking—‘naked statistics’ or evidential pluralism?

Mortality rates differed dramatically between Asian countries which introduced public masking very early in the pandemic and Western countries which did not.¹⁸ But was this relationship causal?

The ‘naked statistics view of evidence-based medicine (EBM)¹² would answer this question exclusively from randomised controlled trial (RCT) evidence, rejecting all other study designs as less trustworthy. The DANMASK researchers, for example, randomised 6024 people to being advised to wear masks outside the home or not.¹⁹ The primary outcome (infection with SARS-CoV-2) occurred in 42 people (1.8%) of the intervention arm and 53 (2.1%) of the control arm—a difference that was not statistically significant. This finding was described by some in the EBM community thus: ‘now that we have properly rigorous scientific research we can rely on, the evidence shows that wearing masks in the community does not significantly reduce the rates of infection’.²⁰

DANMASK had numerous design flaws²¹: it was underpowered; it focused only on whether masks protect the wearer rather than whether they reduce overall community transmission; it occurred at a time of extremely low incidence of COVID-19; it had an intervention period of only 1 month; and—as the authors themselves observed¹⁹—the 95% CIs were compatible with a 46% reduction to a 23% increase in infection with masking (ie, it was inconclusive rather than negative). An editorial in the Cochrane Database of Systematic Reviews encouraged policy-makers to adopt the precautionary principle and ‘act on incomplete evidence’ for several reasons: the urgency of the situation (and the risks of doing nothing), the multiple complexities and complex chains of causation for this particular behavioural intervention, and the potential benefits, over time, of even a non-statistically significant reduction in transmission rate.²²

An EBM+ approach would consider the DANMASK trial and other RCTs (such as a large community masking study in Bangladesh villages²³ on their merits, but also take account of studies which throw light on the mechanism by which masking might work—including.

- ⇒ Laboratory, real-world case studies and mathematical modelling studies supporting an airborne route of transmission.²⁹
- ⇒ Engineering studies which established the filtration properties of different kinds of masks.^{24 25}
- ⇒ Studies on the psychological, sociocultural and wider structural causes of whether people actually wear masks.²⁶
- ⇒ Studies confirming that masking does not appear to have other unintended psychosocial consequences such as risk compensation²⁷ or increased face touching.²⁸

New engineering problems are addressed largely by creatively applying and actively improving existing standards and models. Designing medical equipment such as ventilators or diagnostic imaging systems does not begin de novo but with a standard selection of component products built and tested to meet established standards. Such devices are not iteratively constructed and tested to confirm success. Most aspects are determined analytically using numerical models and design tools. Designs are constrained to meet established national standards that dictate performance and safety requirements. RCTs have no place in such design work.

Engineering knowledge rarely generates a single ‘truth’. Rather, it tends to produce a range of options which vary in how well they meet social needs and satisfy design and resource constraints. There is usually more than one way of designing a medical device to meet safety and performance standards. The success of the engineering method is illustrated by the rarity of harmful failures of engineered systems. When rare failures occur, there follows a detailed investigation of causes, and application of that knowledge to future designs.⁵⁰

Building engineering contributes crucial evidence on how to protect people from airborne pathogens, since the use and impact of personal protective equipment is influenced by (among other things) the quality of indoor air.³⁵ Furthermore, the mechanisms by which filtering facepiece respirators work are long-established and well-understood. Robust certification systems, standards and workplace usage protocols for respirators currently minimise exposure to occupational hazards for millions of workers worldwide. Nobody would propose an RCT comparing these products with less effective protection to ‘prove’ their value in protecting against chemical contamination in a lead smelter—yet such studies are currently being proposed for determining whether respirators protect against biological hazards in healthcare environments.

Human mechanisms: social and behavioural evidence

The human (or social) sciences seek to explain why people act in the way they do, taking account of wider influences such as cultural norms or economic constraints. The Medical Research Council’s framework on developing and evaluating complex interventions emphasises the need for developing a programme theory that takes account of the many influences on human (and organisational and system) behaviour.³⁹

One such framework is the capability-opportunity-motivation-behaviour model,⁵¹ whose lead developer, Professor Susan Michie, was an adviser to UK’s Scientific Advisory Group on Emergencies. In relation to wearing masks or respirators, people may not be physically able to do so, or they may lack key knowledge (eg, they may wrongly consider themselves at low risk, wrongly believe that cloth masks, surgical masks and respirators are equally effective, or underestimate the importance of a good fit); because they lack motivation (including conscious beliefs and subconscious emotional reactions), or because they lack opportunity (eg, their employer bans masks or social pressures convey negative norms).

Other frameworks, such as those based on social ecology theory, place greater emphasis on the multiple layers of influence which shape and constrain human behaviour (including individual, interpersonal, organisational, community and public policy).⁵²

Without a careful exploration of these wider influences, experimental studies of complex interventions may produce misleading findings.

Table 4 Traditional biomedical versus complex systems paradigms (reproduced and adapted under CC-BY-4.0 licence from Greenhalgh and Papoutsis)⁴¹

	Traditional biomedical paradigm	Complex systems paradigm
Scientific truth viewed as	Singular (ie, there is one set of facts which will be found by using the correct scientific methods)	Multiple (ie, there are many legitimate scientific perspectives on a problem, so multiple methods may be needed)
Goal of research	Establishing the truth, which is seen as universal and generalisable	Exploring tensions; generating insights; revealing multiple perspectives
Assumed model of causality	Linear, cause-and-effect; a fixed input has a fixed effect size	Emergent: many interacting influences, but none has a fixed 'effect size'
Typical format of research question	'What is the effect size of the intervention on the predefined outcome, and is it statistically significant?'	'What combination of influences has generated this phenomenon? What does the intervention of interest contribute? What happens to the system and its actors if we intervene in a particular way? What are the unintended consequences elsewhere in the system?'
Good research defined in terms of	Methodological rigour, standardisation, precision	Strong theory, flexible methods, pragmatic adaptation to changing circumstances
Most valued study designs	Randomised controlled trials, meta-analyses	No single design is 'most valued'. A combination of designs—for example, multilevel intervention studies, naturalistic case studies, modelling studies—is preferred
Purpose of theorising	Disjunctive (simplification and abstraction)	Conjunctive (drawing together different parts of a complex problem)
Approach to data collection	Continue collecting data until the dataset is complete and agreed	Data will never be complete; decisions must sometimes be made on the basis of incomplete or contested data
Analytic focus	Dualisms: comparing A with B; finding the influence of X on Y	Dualities: inter-relationships and tensions between A, B, C and other emergent phenomena

Conclusion

The pandemic brought complex and fast-changing problems. We have argued that EBM's traditional methods and quality standards, which favour probabilistic evidence from clinical trials, must be extended to place greater emphasis on other sources of evidence. We are not, however, arguing for an anything-goes approach to evidence or what one reviewer of an earlier draft of this paper called 'a throwback to the 1970s'—a time before much of the rigorous scholarship of EBM had been built.

An urgent debate is needed within the EBM/evidence-based healthcare community on how, in what circumstances and to what extent EBM should evolve into EBM+. We hope to contribute to that debate with further articles on how an EBM +approach can enhance and extend EBM's important contribution to pandemic science.

Twitter Trisha Greenhalgh @trishgreenhalgh and Matthew Oliver @sameo416

Acknowledgements Thanks to Kevin Hedges, Leyla Asadi, and Simon Smith and a colleague who preferred to remain anonymous for helpful comments on previous versions of this paper.

Contributors TG wrote the original draft. All authors contributed to planning, discussing, drafting and refining the manuscript.

Funding TG's work is partly supported by the NIHR Oxford Biomedical Research Centre (BRC-1215-20008). DC received no specific grant from any funding agency in the public, commercial or not-for-profit sectors. MO received no specific grant from any funding agency in the public, commercial or not-for-profit sectors. RM is funded by the National Health and Medical Research Council Principal Research Fellowship (grant number 1137582). DF is supported by Canadian Institutes for Health Research (2019 COVID-19 rapid researching funding OV4-170360).

HYPOTHESIS TESTING IN CONTROLLED EXPERIMENTS (e.g. TRIALS)

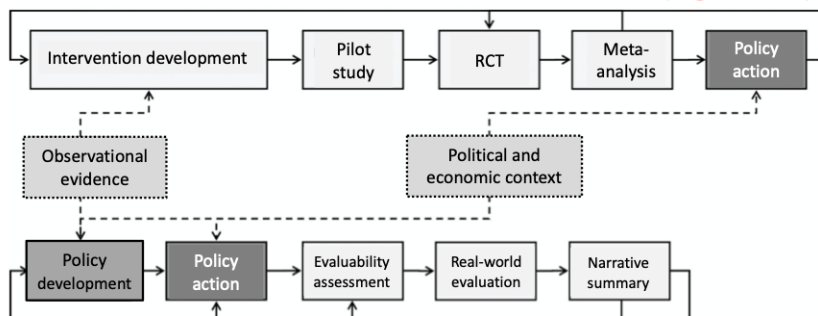


Figure 2 Combining controlled trials with evaluation of real-world action via case studies. (Reproduced and adapted under CC-BY-4.0 Licence from Ogilvie *et al*⁴³). RCT, randomised controlled trial.

Competing interests TG is a member of Independent SAGE and an unpaid adviser to Balvi, a philanthropic fund. RM in the last 5 years has received funding from Sanofi and Seqirus for investigator driven research on influenza vaccines. She has been on advisory boards or consulted for COVID-19 vaccines for Seqirus, Janssen and Astra-Zeneca. She has received funding for an industry-linkage grant scheme from mask manufacturer Detmold and consulted for mask manufacturer Ascend Performance Technologies. DF has served on advisory boards related to influenza and SARS-CoV-2 vaccines for Seqirus, Pfizer, Astrazeneca and Sanofi-Pasteur Vaccines, and has served as a legal expert on issues related to COVID-19 epidemiology for the Elementary Teachers Federation of Ontario and the Registered Nurses Association of Ontario. DJC worked as a Clinical Specialist for the Canadian PPE Manufacturers Association between November 2021 and February 2022. She now serves as an executive board member with the Coalition of Healthcare Associated Infection Reduction (CHAIR), a volunteer, not-for-profit group, and is a co-founder of the volunteer grassroots group Community Access to Ventilation Information (CAVI) supporting public library CO2 monitor loaning programmes.

Patient consent for publication Not applicable.

Ethics approval Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

Trisha Greenhalgh <http://orcid.org/0000-0003-2369-8088>

Matthew Oliver <http://orcid.org/0000-0002-7197-3032>

References

- Tresker S. Treatment Effectiveness and the Russo-Williamson Thesis, EBM+, and Bradford Hill's Viewpoints. *Int Stud Philos Sci* 2022;33:1-28.
- Aronson JK, Auker-Howlett D, Ghiara V, et al. The use of mechanistic Reasoning in assessing coronavirus interventions. *J Eval Clin Pract* 2021;27:684-93.
- Parkkinen V-P, Wallmann C, Wilde M. *Evaluating evidence of mechanisms in medicine: principles and procedures*. Springer Nature, 2018.
- Murad MH, Asi N, Alsawas M, et al. New evidence pyramid. *Evid Based Med* 2016;21:125-7.
- Guyatt G, Cairns J, Churchill D, et al. Evidence-Based medicine. A new approach to teaching the practice of medicine. *JAMA* 1992;268:2420-5.
- Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008;336:924-6.
- Howick J, Chalmers I, Glasziou P. Explanation of the 2011 Oxford centre for evidence-based medicine (OCEBM) levels of evidence (background document) (2011) Oxford Centre for Evidence-Based Medicine; 2018. <http://www.cebm.net>
- RECOVERY Collaborative Group, Horby P, Lim WS, et al. Dexamethasone in hospitalized patients with Covid-19. *N Engl J Med* 2021;384:693-704.
- Butler CC, Dorward J, Yu L-M, et al. Azithromycin for community treatment of suspected COVID-19 in people at increased risk of an adverse clinical course in the UK (PRINCIPLE): a randomised, controlled, open-label, adaptive platform trial. *Lancet* 2021;397:1063-74.
- PANORAMIC trial group. PANORAMIC: platform adaptive trial of novel antivirals for early treatment of COVID-19 in the Community Oxford: Nuffield department of primary care health sciences. Available: <https://www.panoramictrial.org/for-healthcare-professionals/documents> 2022 [Accessed 19 Jun 2022].
- Khalil H, Tamara L, Rada G, et al. Challenges of evidence synthesis during the 2020 COVID pandemic: a scoping review. *J Clin Epidemiol* 2022;142:10-18.
- Clarke B, Gillies D, Illari P, et al. Mechanisms and the evidence hierarchy. *Topoi* 2014;33:339-60.
- Williamson J. Establishing causal claims in medicine. *International Studies in the Philosophy of Science* 2019;32:33-61.
- Williamson J. Establishing the teratogenicity of Zika and evaluating causal criteria. *Synthese* 2021;198:2505-18.
- HILL AB, Bradford Hill A. The environment and disease: association or causation? *Proc R Soc Med* 1965;58:295-300.
- Russo F, Williamson J. Interpreting causality in the health sciences. *International Studies in the Philosophy of Science* 2007;21:157-70.
- Fedak KM, Bernal A, Capshaw ZA, et al. Applying the Bradford Hill criteria in the 21st century: how data integration has changed causal inference in molecular epidemiology. *Emerg Themes Epidemiol* 2015;12:1-9.
- Leffler CT, Ing E, Lykins JD, et al. Association of country-wide coronavirus mortality with demographics, testing, lockdowns, and public wearing of masks. *Am J Trop Med Hyg* 2020;103:2400-11.
- Bundgaard H, Bundgaard JS, Raaschou-Pedersen DET, et al. Effectiveness of Adding a Mask Recommendation to Other Public Health Measures to Prevent SARS-CoV-2 Infection in Danish Mask Wearers: A Randomized Controlled Trial. *Ann Intern Med* 2021;174:335-43.
- Heneghan C, Jefferson T. Landmark Danish study finds no significant effect for facemask wearers. *The Spectator* 2020.
- Frieden TR, Cash-Goldwasser S. Of masks and methods. *Ann Intern Med* 2021;174:421-2.
- Soares-Weiser K, Lasserson T, Jorgensen KJ, et al. Policy makers must act on incomplete evidence in responding to COVID-19. *Cochrane Database Syst Rev* 2020;11:ED000149.
- Abaluck J, Kwong LH, Styczynski A, et al. Impact of community masking on COVID-19: a cluster-randomized trial in Bangladesh. *Science* 2022;375:eabi9069.
- Konda A, Prakash A, Moss GA. Aerosol filtration efficiency of common fabrics used in respiratory cloth masks. *ACS Nano* 2020.
- Hill WC, Hull MS, MacCuspie RI. Testing of commercial masks and respirators and cotton mask insert materials using SARS-CoV-2 virion-sized particulates: comparison of ideal aerosol filtration efficiency versus fitted filtration efficiency. *Nano Lett* 2020;20:7642-7.
- van der Westhuizen H-M, Kotze K, Tonkin-Crine S, et al. Face coverings for COVID-19: from medical intervention to social practice. *BMJ* 2020;370:e3020.
- Mantzari E, Rubin GJ, Marteau TM. Is risk compensation threatening public health in the covid-19 pandemic? *BMJ* 2020;370:m2913.
- Chen Y-J, Qin G, Chen J, et al. Comparison of Face-Touching behaviors before and during the coronavirus disease 2019 pandemic. *JAMA Netw Open* 2020;3:e2016924-e24.
- Greenhalgh T, Jimenez JL, Prather KA, et al. Ten scientific reasons in support of airborne transmission of SARS-CoV-2. *Lancet* 2021;397:1603-5.
- Djulfbegovic B, Guyatt G. Evidence-Based medicine in times of crisis. *J Clin Epidemiol* 2020;126:164-6.
- Jefferson T, Del Mar CB, Dooley L, et al. Physical interventions to interrupt or reduce the spread of respiratory viruses. *Cochrane Database Syst Rev* 2020;11:CD006207.
- van Doremalen N, Bushmaker T, Morris DH, et al. Aerosol and surface stability of SARS-CoV-2 as compared with SARS-CoV-1. *N Engl J Med* 2020;382:1564-7.
- Kutter JS, de Meulder D, Bestebroer TM, et al. SARS-CoV and SARS-CoV-2 are transmitted through the air between ferrets over more than one meter distance. *Nat Commun* 2021;12:1-8.
- Chen W, Zhang N, Wei J, et al. Short-Range airborne route dominates exposure of respiratory infection during close contact. *Build Environ* 2020;176:106859.

- 35 Morawska L, Allen J, Bahnfleth W, *et al.* A paradigm shift to combat indoor respiratory infection. *Science* 2021;372:689–91.
- 36 Miller SL, Nazaroff WW, Jimenez JL, *et al.* Transmission of SARS-CoV-2 by inhalation of respiratory aerosol in the Skagit Valley Chorale superspreading event. *Indoor Air* 2021;31:314–23.
- 37 Wang CC, Prather KA, Sznitman J, *et al.* Airborne transmission of respiratory viruses. *Science* 2021;373:eabd9149.
- 38 Campbell M, Fitzpatrick R, Haines A, *et al.* Framework for design and evaluation of complex interventions to improve health. *BMJ* 2000;321:694–6.
- 39 Skivington K, Matthews L, Simpson SA, *et al.* A new framework for developing and evaluating complex interventions: update of medical Research Council guidance. *BMJ* 2021;374:n2061.
- 40 Borgerson K. Valuing evidence: bias and the evidence hierarchy of evidence-based medicine. *Perspect Biol Med* 2009;52:218–33.
- 41 Greenhalgh T, Papoutsis C. Studying complexity in health services research: desperately seeking an overdue paradigm shift. *BMC Med* 2018;16:95.
- 42 Papparini S, Papoutsis C, Murdoch J. Evaluating complex interventions in context: systematic, meta-narrative review of case study approaches. *BMC Medical Research Methodology* 2021;21:1–22.
- 43 Ogilvie D, Adams J, Bauman A, *et al.* Using natural experimental studies to guide public health action: turning the evidence-based medicine paradigm on its head. *J Epidemiol Community Health* 2020;74:203–8.
- 44 Greenhalgh T, Thorne S, Malterud K. Time to challenge the spurious hierarchy of systematic over narrative reviews? *Eur J Clin Invest* 2018;48:e12931.
- 45 Ogilvie D, Bauman A, Foley L, *et al.* Making sense of the evidence in population health intervention research: building a dry stone wall. *BMJ Glob Health* 2020;5:e004017.
- 46 Pandemic Influenza Outbreak Research Modelling Team (Pan-InfORM), Fisman D. Modelling an influenza pandemic: a guide for the perplexed. *CMAJ* 2009;181:171–3.
- 47 Flyvbjerg B. Five misunderstandings about case-study research. *Qualitative Inquiry* 2006;12:219–45.
- 48 Koen BV. Toward a definition of the engineering method. *European Journal of Engineering Education* 1988;13:307–15.
- 49 American Society of Heating R, and Air-Conditioning Engineers. ANSI/ASHRAE/ASHE standard 170-2017, ventilation of health care facilities. Washington, DC ASHRAE; 2022. <https://www.ashrae.org/technical-resources/standards-and-guidelines/standards-addenda/ansi-ashrae-ashrae-standard-170-2017-ventilation-of-health-care-facilities> 2017 [Accessed 17 Feb 2022].
- 50 Englehardt E, Werhane PH, Newton LH. Leadership, engineering and ethical clashes at Boeing. *Sci Eng Ethics* 2021;27:1–17.
- 51 McDonagh LK, Saunders JM, Cassell J, *et al.* Application of the COM-B model to barriers and facilitators to Chlamydia testing in general practice for young people and primary care practitioners: a systematic review. *Implementation Science* 2018;13:1–19.
- 52 Glanz K, Bishop DB. The role of behavioral science theory in development and implementation of public health interventions. *Annu Rev Public Health* 2010;31:399–418.