Book of Abstracts

Causality and Causal Inference in Medicine

Sevilla, October 7, 2026 - October 9, 2026

16th Causality in the Sciences (CitS) Conference series

Grupo de Lógica Lenguaje e Información (HUM609)
Departamento de Filosofía, Lógica y Filosofía de la ciencia
Universidad de Sevilla

Proyecto PID2024-157876NA-I00
Ayuda a grupos de investigación VII PPIT–US 2026/2026/00000336

Organising committee

- Cristina Barés Gómez, Universidad de Sevilla
- Matthieu Fontaine, Universidad de Sevilla
- Elena Popa, Universidad de Sevilla
- Quentin Ruyant, Universidad de Sevilla

Steering committee

- Phyllis Illari, Science & Technology Studies, UCL
- Samantha Kleinberg, Computer Science, Stevens
- Bert Leuridan, Philosophy, Antwerp
- Julian Reiss, Philosophy, Linz
- Federica Russo, History and Philosophy of Science, Utrecht
- Erik Weber, Philosophy, Ghent
- Jon Williamson, Philosophy, Manchester.

Contents

Keynote Sessions

The problems of extrapolation in epidemiological medicine

Jonathan Fuller
University of Pittsburgh

In this talk, I distinguish different components of the problem of extrapolating causal conclusions from health sciences research. The standard approach to making therapeutic predictions in contemporary 'epidemiological medicine' involving extrapolating from the results of epidemiological studies such as clinical trials. I explore the limitations of existing solutions to the problems of extrapolation proposed by philosophers and medical experts. These solutions sometimes conflate the epistemological problem of extrapolation with the pragmatics of extrapolating, or the problem of extrapolating from one population to another with the problem of individualizing population results for single cases. Existing solutions also often overlook distinct logical and evidential components of the problem. I propose a novel solution to the logical problem of extrapolation I call 'mechanism-cofactor analysis' and a novel solution to the evidential problem relying on structural analogies.

Value-ladenness of causal evidence: the case of testing interventions for Alcohol Use Disorder

Saana Jukola
University of Twente

Excessive alcohol use is a major cause of mortality and morbidity, and research-based interventions targeting problematic drinking are needed. The so-called evidence-based approach, according to which decisions about the treatment of patients in clinical and social care context should be based on "conscientious, explicit, and judicious use of current best evidence" (Sackett et al. 1996, 71) has been adopted in the management of alcohol-related diseases since 1990s. According to the tenets of evidence-based medicine and practice (EBP), randomized controlled trials and meta-analyses are the best means for securing the objectivity of results. A reason for this is that they are believed to limit the need for making potentially value-laden judgments during the process. Despite the criticism this position has received (e.g., Truijens 2017), EBP guidelines govern the availability of both pharmaceutical and non-medication treatment options as well as the allocation of resources in many countries (Klingeman & Storbjörk 2016).

By discussing how the effectiveness of interventions for Alcohol Use Disorder (AUD) is tested, this talk addresses conceptual and epistemological issues that challenge some of the assumptions that underlie EBP. For example, I show that researchers rely on value-laden premises concerning the disease definition and diagnostic criteria when designing trials. Dealing with the value-ladenness inherent in the construction of trials precludes the possibility of achieving objectivity in the sense emphasized by the promoters of EBP. Consequently, there is an inherent tension in testing interventions for AUD within the EBP framework.

References

Klingemann, H., & Storbjörk, J. (2016). The treatment response: Systemic features, paradigms and socio-cultural frameworks. The SAGE handbook of drug and alcohol studies: Social science approaches, 1, 260-286.

Sackett, D. L., Rosenberg, W. M., Gray, J. M., Haynes, R. B., & Richardson, W. S. (1996). Evidence based medicine: what it is and what it isn't. bmj, 312(7023), 71-72.

Truijens, F. L. (2017). Do the numbers speak for themselves? A critical analysis of procedural objectivity in psychotherapeutic efficacy research. Synthese, 194(12), 4721-4740.

Beyond Single Causes: A Complexity Perspective on Causal Inference in Public Health

Naja Hulvej Rod
University of Copenhagen

Modern causal inference has transformed medicine and epidemiology by providing rigorous methods for estimating the effects of interventions. Yet many of today's greatest public health challenges, including health inequalities, mental health, multimorbidity, and chronic disease, cannot be fully understood through single-cause frameworks alone. These problems emerge from complex systems in which biological, social, behavioral, and environmental factors interact across time and scales. Embracing this complexity does not mean abandoning causal inference; rather, it means expanding the range of causal questions we ask and the methods we use to answer them.

This talk will present a complexity perspective on public health that integrates traditional epidemiologic methods with systems thinking, computational approaches, and insights from philosophy and the social sciences. This will include a discussion on how a pluralistic approach to causality can strengthen our understanding of health by combining evidence on patterns, mechanisms, and system dynamics, while highlighting why interdisciplinary collaboration is essential for understanding complex health problems. Ultimately, the future of causal research in public health lies not in choosing between causal inference and complexity science, but in integrating them to better understand and improve the health of populations.

Ordinary Sessions

Beyond Molecules: Psychological Variables and Causal Inference in Randomized Controlled Trials

Héloïse Athea
CNRS / Université Paris 1

Randomized controlled trials (RCTs) are classically designed to isolate the causal effect of a pharmacological intervention by controlling for potential confounders through randomization and standardized protocols. An implicit conception of biomedical causality typically underlies this framework, according to which therapeutic efficacy is primarily attributable to the biochemical action of a molecule on underlying pathophysiological processes.

However, a growing body of research suggests that psychological variables - such as patient expectations, therapeutic alliance, emotional support, and broader care conditions - can significantly influence clinical trajectories, including symptom burden, treatment adherence, quality of life, and even (more controversially) disease progression in severe somatic illnesses such as cancer. These effects are mediated by complex, multilevel mechanisms involving neuroendocrine, immune, and inflammatory pathways. This challenges the assumption, implicit in standard RCT designs, that a clear separation can be drawn between pharmacological and contextual or psychological factors in causal inference.

In this context, this presentation addresses the following question: should psychological and contextual variables be treated as confounders to be controlled for, or as mediators or constitutive components of the therapeutic intervention itself that require systematic characterization? How does this choice reflect underlying epistemological commitments regarding what counts as a cause, what counts as good evidence, and how an "intervention" ought to be individuated (i.e., whether it includes only the molecule or also its delivery context) within clinical research?

By examining these issues in the context of oncology, this talk aims to make explicit a set of often implicit epistemic and methodological assumptions embedded in RCTs, and to assess their implications for causal inference in contemporary clinical research.

References

Cartwright, N. (2007). Hunting Causes and Using Them. Cambridge University Press.

Cartwright, N., & Munro, E. (2010). "The limitations of randomized controlled trials in predicting effectiveness." Journal of Evaluation in Clinical Practice, 16(2), 260–266.

Fu, W. W., Popovic, M., Agarwal, A., Milakovic, M., Fu, T. S., McDonald, R., … &

Chow, E. (2016). The impact of psychosocial intervention on survival in cancer: a meta-analysis. Annals of Palliative Medicine, 5(2), 9306-9106.

Howick, J. (2011). The Philosophy of Evidence-Based Medicine. Wiley-Blackwell.

Illari, P. M., & Williamson, J. (2012). "What is a mechanism?" European Journal for Philosophy of Science, 2(1), 119–135.

Kiecolt-Glaser, Janice K., et al. "Psychoneuroimmunology and psychosomatic medicine: back to the future." Biopsychosocial Science and Medicine 64.1 (2002): 15-28.

Reiss, J. (2015). Causation, Evidence, and Inference. Routledge.

Russo, F., & Williamson, J. (2007). "Interpreting causality in the health sciences." International Studies in the Philosophy of Science, 21(2), 157–170.

Spiegel, D., Kraemer, H., Bloom, J., & Gottheil, E. (1989). Effect of psychosocial treatment on survival of patients with metastatic breast cancer. The lancet, 334(8668), 888-891.

Spiegel, D., Butler, L.D., Giese-Davis, J., Koopman, C., Miller, E., DiMiceli, S., Classen, C.C., Fobair, P., Carlson, R.W. and Kraemer, H.C. (2007), Effects of supportive-expressive group therapy on survival of patients with metastatic breast cancer . Cancer, 110: 1130-1138.

Temoshok, L. R., & Wald, R. L. (2002). Change is complex: Rethinking research on psychosocial interventions and cancer. Integrative Cancer Therapies, 1(2), 135-145.

Worrall, J. (2007). "Evidence in medicine and evidence-based medicine." Philosophy Compass, 2(6), 981–1022.

Beyond the Epistemic Loop – Mechanistic Hypotheses in Biomedical Reasoning

Cristina Barés Gómez
University of Seville

Matthieu Fontaine
University of Seville

Biomedical practice and inquiry are fundamentally causal; physicians intervene in diseases causes to affect patient well-being, while researchers seek causal links between treatments and effects. However, since causal links cannot be directly observed, they are inferred through statistical correlations. Within the Evidence-Based Medicine (EBM) paradigm (EBM-Group 1992), this has led to hierarchies where Randomized Controlled Trials (RCTs) are placed at the top, while mechanisms are often dismissed as speculative and relegated to the bottom.

The Russo-Williamson thesis (2007) argues that both difference-making studies and mechanistic evidence are necessary to establish causal claims. Yet, a conceptual challenge remains: if "mechanistic evidence" is established through empirical evidence of entities and activities (Illari, 2011), it becomes difficult to distinguish it from standard of EBM.

We propose approaching this issue from an inferentialist perspective, analysing the role of mechanism in medical reasoning–a complex process that involves the Peircean triad of abduction, deduction, and induction. A central point of our argument is the rejection of abduction as Inference to the Best Explanation (IBE) (Harman 1965). Defining abduction through "explanation" is inherently circular: it puts the "explanatory cart" before the "abductive horse" by assuming a classification of the "best" hypothesis is already available, thereby conflating abduction with induction. Furthermore, abduction is often not even explanatory at all, but rather a process in the course of which an "interrogation" or an "ignorance-preserving" hypothesis that initiates the inquiry (Hintikka 1998, Gabbay and Woods 2005).

Drawing on Magnani's Select and Test Model (1992) and Gabbay and Woods Model of abduction (2005), and recent developments of Barés and Fontaine (2025), we distinguish two general forms of reasoning, in the context of which the use of mechanistic evidence is clarified:

1. Inference to mechanisms: evidence used to corroborate mechanistic hypotheses;

2. Inference from mechanisms: using established or conjectured mechanisms as "reason to accept" a biomedical or a clinical claim-

Our point is that EBM cannot function as an isolated epistemic system because it inherently assumes a prior identification of which variables and data are relevant to the inquiry. Identifying what to "look for" in clinical trials necessitates a mechanistic

hypothesis to guide the initial selection of data. This mirrors the fundamental limitation of "naked" Baconian induction, already stressed by authors such as Peirce (1931/1958) or Hempel (1966): without an abductive hypothesis to organize the manifold of observation, induction lacks a principle for gathering and refining information. By positioning the mechanism as the "initial guess" that initiates investigation, rather than an IBE-style conclusion, we resolve an apparent circularity involving EBM and mechanisms.

References

Barés Gómez, C., & Fontaine, M. 2025. Mechanisms, Evidence, and Abductive Hypotheses. Global Philosophy 35(8).

Evidence-Based Medicine Working Group. 1992. Evidence-based medicine. A new approach to teaching the practice of medicine. JAMA: 268(17): 2420-5.

Gabbay, D., & Woods, J. 2005. The Reach of Abduction. Insight and Trials. Amsterdam. Elsevier.

Harman, G. H. 1965. The Inference to the Best Explanation. The Philosophical Review 74(1): 88-95.

Hempel, C.G. 1966. Philosophy of Natural Science. London. Prentice-Hall.

Hintikka, J. 1998. What is Abduction? The Fundamental Problem of Contemporary Epistemology. Transactions of the Charles S. Peirce Society 34(3): 503-534.

Illari, P. 2011. Mechanistic evidence: Disambiguating the Russo-Williamson thesis. International Studies in the Philosophy of Science 25(2): 139–157.

Magnani, L. 1992. Abductive Reasoning: Philosophical and Educational Perspectives in Medicine. In Evans, D. et al. (eds.), Advanced Models of Cognition for Medical Training and Practice. Berlin. Springer: 21-41.

Peirce, C.S. 1937/1958. Collected Papers of Charles Sanders Peirce. Harvard. HUP.

Russo, F., & Williamson, J. 2007. Interpreting causality in the health sciences. International Studies in the Philosophy of Science 21(2): 157–170.

Managing treatment selection in oncology: the epistemic rationale of risk stratification

Marina Antonia D'Amico
Scuola Universitaria Superiore IUSS Pavia

Giovanni Sanavio
Scuola Universitaria Superiore IUSS Pavia

The growing reliance on data-driven tools for personalized medicine, particularly in oncology, raises the question of what methods and conceptual approaches could provide a reliable basis for tailoring treatments to individual patients. Risk stratification, in this context, represents an essential instrument for determining individual prognoses and guiding treatment selection. It involves estimating the likelihood of adverse clinical outcomes based on patient-specific parameters, like biomarkers and other clinical variables identified through retrospective studies, and assigning patients to low- or high-risk groups. Risk stratification allows a comparison of the effectiveness of different therapeutic strategies, thereby helping to optimize resource utilization (Vogt, 2025).

Our contribution aims to evaluate the epistemic rationale of risk stratification in medical reasoning. To do so, we analyze risk stratification as an instance of the reference class problem: stratifying patients according to risk involves assigning them to a class of individuals sharing homogeneous relevant characteristics (Hájek, 2007; Devanesan, 2023; Fuller and Flores, 2015). The key issue is thus how to justify which features are relevant for determining risk exposure—that is, for establishing an individual's membership to a given reference class. In other words, what kind of evidence supports attributing a certain chance of an adverse outcome to a particular individual?

To answer this question, we explore whether the reference class problem—which arises in diagnosis, and disease occurrence prediction—exhibits unique features in the case of treatment selection. Building on the idea that the choice of the appropriate reference class depends on the purpose the reasoning is intended to serve (Parker, 2020), we explore how treatment selection imposes *distinctive* constraints on justifying relevant data and classes. Drawing on Wallmann and Williamson's (2017) claim that the relevant reference class is constrained by a causal structure, we argue that treatment selection imposes distinctive *causal* constraints on reference class justification. Using Pearl's interventionist causal framework (2009), we show that treatment selection involves a different causal structure than diagnosis or disease occurrence prediction—and thus imposes peculiar constraints on reference classes in risk stratification.

These ideas will be tested against a medical case study involving the International Metastatic Disease-Stage Classification (IMDC) and Memorial Sloan Kettering Cancer Center (MSKCC) risk models for metastatic renal cell carcinoma (mRCC) (Heng et al., 2009; Méjean et al., 2018; Motzer et al., 2001; Okita et al., 2019; Keskinkilic et al., 2025). We will illustrate how causal constraints account for the superior performance of

the IMDC model compared to the MSKCC model in predicting patient survival across risk groups.

Our account of risk stratification has the potential to contribute to the refinement of decision-making practices for personalized treatment selection, even in cases where statistical evidence is not readily available, offering a rigorous framework for robust clinical reasoning.

References

Devanesan, A. (2023). The reference class problem and probabilities in the individual case: A response to Fuller. Philosophy of Science, 90(4), 1001-1009.

Fuller, J., & Flores, L. J. (2015). The Risk GP Model: The standard model of prediction in medicine. Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, 54, 49-61.

Hájek, A. (2007). The reference class problem is your problem too. Synthese, 156(3), 563-585.

Heng, D. Y., Xie, W., Regan, M. M., Warren, M. A., Golshayan, A. R., Sahi, C., … & Choueiri, T. K. (2009). Prognostic factors for overall survival in patients with metastatic renal cell carcinoma treated with vascular endothelial growth factor–targeted agents: results from a large, multicenter study. Journal of clinical oncology, 27(34), 5794-5799.

Keskinkılıç, M., Canaslan, K. Ü. B. R. A., Semiz, H. Ü. S. E. Y. İ. N., & Yavuzşen, T. U. Ğ. B. A. (2025). Comparison of Prognostic Risk Models (IMDC, MSKCC, CFF) in Patients Diagnosed with Metastatic Renal Cell Cancer. Forbes Journal of Medicine.

Méjean, A., Ravaud, A., Thezenas, S., Colas, S., Beauval, J. B., Bensalah, K., … & Escudier, B. (2018). Sunitinib alone or after nephrectomy in metastatic renal-cell carcinoma. New England Journal of Medicine, 379(5), 417-427.

Motzer, R. J., Bacik, J., Murphy, B. A., Russo, P., & Mazumdar, M. (2002). Interferon-alfa as a comparative treatment for clinical trials of new therapies against advanced renal cell carcinoma. Journal of clinical oncology, 20(1), 289-296.

Okita, K., Hatakeyama, S., Tanaka, T., Ikehata, Y., Tanaka, T., Fujita, N., … & Ohyama, C. (2019). Impact of disagreement between two risk group models on prognosis in patients with metastatic renal-cell carcinoma. Clinical Genitourinary Cancer, 17(3), e440-e446.

Parker, W. S. (2020). Model evaluation: An adequacy-for-purpose view. Philosophy of Science, 87(3), 457-477.

Pearl, J. (2009). Causality. Cambridge university press.

Vogt, H. (2025). Personalized medicine beyond stratification. In Handbook of the Philosophy of Medicine (pp. 1241-1263). Dordrecht: Springer Netherlands.

Wallmann, C., & Williamson, J. (2017). Four approaches to the reference class problem. In Making it formally explicit: probability, causality and indeterminism (pp. 61-81). Cham: Springer International Publishing.

Machine learning for philosophy-based causal discovery

Luna De Souter
Witten/Herdecke University

Coincidence Analysis (CNA) is a method for automatically discovering complex causal structures from datasets. When applied to a dataset, this method algorithmically builds models that satisfy the criteria of causation spelled out by the philosophical regularity theory of Baumgartner and Falk (2023) for that data. CNA has been applied for causal discovery across scientific disciplines, and is most frequently used in healthcare research, for instance, to study opioid treatment (Matson et al. 2025), cancer screening outreach (Coury et al. 2021), and surgical site infection reduction (Schlick et al. 2022).

A central limitation to CNA's applicability in healthcare research is the method's inability to analyze datasets comprising more than 30 variables. Since many healthcare datasets are considerably larger, researchers must often eliminate variables before applying CNA, which may jeopardize the reliability of their results.

In this presentation, I argue that the Boolean Rule Column Generation (BRCG) algorithm (Dash, Günlük, and Wei 2018), which was originally developed by machine learning researchers as an algorithm for building interpretable binary classification models for datasets comprising hundreds of variables, can and should also be used as an algorithm for discovering complex causal structures from datasets with more variables than CNA is able to handle.

My main argument is conceptual: I demonstrate that, if the BRCG algorithm succeeds at optimizing its objective function, it returns a model that best satisfies the criteria of Baumgartner and Falk's (2023) theory for the analyzed data. Since BRCG is not guaranteed to optimize its objective function, I complement this conceptual argument with simulation studies showing that BRCG performs similarly to CNA on datasets small enough for CNA to handle, while reliably recovering causal structures from datasets with more variables than CNA can handle. Finally, I address the objection that BRCG's opacity makes it less trustworthy than CNA by arguing that, in causal discovery for healthcare research, the benefits of avoiding the elimination of confounding variables outweigh the costs of increased opacity.

References

M. Baumgartner and C. Falk. Boolean difference-making: A modern regularity theory of causation. 2023. The British Journal for the Philosophy of Science, 74(1): 171-197. doi: 10.1093/bjps/axz047.

S. Dash, O. Günlük, and D. Wei. 2018. Boolean decision rules via column

generation. In Proceedings of the 32nd International Conference on Neural Information Processing Systems. Curran Associates Inc., New York, 4660-4670. url: https://dl.acm.org/doi/pdf/10.5555/3327345.3327376.

T. Matson, A. Lee, E. Miech, P. Wartko, R. Phillps, M. Shea, et al. 2021. The difference-making rule of staff support in implementing nurse care management for opioid use disorder treatment: A configurational analysis. Journal of Substance Use and Addiction Treatment. doi: 10.1016/j.josat.2025.209642.

J. Coury, E. Miech, P. Styer, A. Petrik, K. Coates, B. Green, et al. 2021. What's the "secret sauce"? How implementation variation affects the success of colorectal cancer screening outreach. Implementation Science Communications, 2(1). doi: 10.1186/s43058-020-00104-7.

C. Schlick, R. Huang, B. Brajcich, A. Halverson, A. Yang, L. Kreutzer, et al. 2022. Unbundling bundles: Evaluating the association of individual colorectal surgical site infection reduction bundle elements on infection rates in a statewide collaborative. Diseases of the Colon & Rectum. 65(8): 1052-1061. doi: 10.1097/DCR.0000000000002223.

A causal account for estimating effects across populations when no causal information is available

Alexander Gebharter
Marche Polytechnic University
alexander.gebharter@gmail.com

Barbara Osimani
Marche Polytechnic University
barbaraosimani@gmail.com

Michal Sikorski
Marche Polytechnic University
michalpsikorski@gmail.com

Elisa D'Olimpio
Marche Polytechnic University
elisa.dolimpio@studenti.unicam.it

Zhitao Zhang
Marche Polytechnic University
zhangzhitao743@gmail.com

The question of how experimental results obtained in a study population can be used to obtain an unbiased estimate of causal effects in a target population is key in evidence-based medicine. The target population can be the population from which the study population was sampled or a completely different population. In both cases, certain factors may be distributed differently in the two populations. This may affect the causal effect of interest, depending on the specific causal structure and on how the factors to be intervened on interact with the variables whose values are distributed differently (cf. Cartwright & Hardie, 2012). This is generally addressed by assuming factors to be distributed similarly enough among the individuals of the study and the target population. However, in practice this assumption is often violated, information about the causal structure is limited, and heterogeneity is a common problem. In this presentation, we use causal models to develop an account that can overcome this problem.

The presentation is structured into 4 main parts. In part 1, we introduce the formal basics relevant for our endeavour. In part 2, we present and discuss Pearl and Bareinboim's (2011) account of extrapolating experimental results across populations. Their approach is based on so-called selection variables, which are technical means to represent heterogeneity by allowing one to manipulate the distribution of a variable's values among individuals of the study or the target population. While Pearl and Bareinboim aim at an exact computation of the causal effect in the target population, given full

structural knowledge about the study and the target population, we rather focus on the more practical question of how the effect in the target population should be estimated if causal information is limited. We do not see our account as an alternative to Pearl and Bareinboim's, but rather as supplementary. Nevertheless, we point to a problem with their account and highlight a class of relevant causal settings in which their approach delivers counterintuitive results.

In part 3, we diagnose the source of the problem for Pearl and Bareinboim's (2011) approach and develop our own account. The basic idea is to account for varying distributions of causal factors by Jeffrey conditionalization. We demonstrate how the account provides the intuitively correct results for the scenarios in which Pearl and Bareinboim's account fails and propose a transport criterion for identifying cases where experimental results can be used to estimate an intervention's effect in the target population. In part 4, we propose an empirical test for the developed criterion. Although our account uses causal models to generate the general recipe for estimating an intervention's effect in the target population, its application given a positive empirical test does not require causal knowledge about the two domains. The account is, thus, causally warranted without requiring specific causal information.

References

Cartwright, N., & Hardie, J. (2012). Evidence-based policy: A practical guide to doing it better. Oxford University Press.

Pearl, J., & Bareinboim, E. (2011). External validity and transportability: A formal approach. In Proceedings of the Joint Statistical Meetings (JSM) (pp. 157–171). American Statistical Association.

Why Evidence-Based Medicine+ Should Care about Surrogate Outcomes

Kyrellos Ghattas
Northern Illinois University

Philosophical proponents of Evidence-Based Medicine+ argue that mechanistic evidence is necessary for making causal claims in medicine. They further assert that the dominant paradigm in medical literature (i.e. Evidence-Based Medicine) is epistemically impoverished due to its sidelining of mechanistic evidence. My aim in this presentation is to show that there is already a genre of medical literature, the surrogate outcomes literature, that makes use of mechanistic evidence when assessing causal claims between surrogate outcomes and target outcomes. Furthermore, I argue that given the use of mechanistic evidence in the surrogate outcomes literature, proponents of Evidence-Based Medicine+ (EBM+) ought to turn to the surrogate outcomes literature for empirical insights on the use of mechanistic evidence in establishing causal claims in medicine.

My argument bridges EBM+'s guidelines on mechanistic evidence and the surrogate outcomes literature via two claims – the Broad Claim, and the Focused Claim. The Broad Claim highlights how regulatory agencies take mechanistic evidence to be necessary in determining whether surrogate outcomes are causally linked to target outcomes. This is in line with EBM+'s assertion that mechanistic evidence is necessary for establishing causal claims in medicine. The Focused Claim demonstrates how the Prentice Criterion, a biostatistical method of surrogate outcome validation, operationalizes the assessment of mechanistic evidence when evaluating the causal connection between a proposed surrogate outcome and a target outcome. The Focused Claim further asserts that the Prentice Criterion directly accords with the Russo-Williamson thesis (a core doctrine of EBM+).

At the end of my argument, I will have shown that the surrogate outcomes literature uses mechanistic evidence to establish causal claims in a way that is congruent with the recommendations of EBM+. Both EBM+ and the surrogate outcomes literature take mechanistic evidence to be necessary in demonstrating causal claims in medicine. And the Prentice Criterion used in the surrogate outcomes literature is an operationalized method of EBM+'s recommendations for assessing mechanistic evidence.

The implication of this argument is that philosophical proponents of EBM+ have good reasons to explore the surrogate outcomes literature; specifically for insights on the empirical benefits and drawbacks of using mechanistic evidence in establishing causal claims in medicine. In the final section of my presentation, I give a preview of what this process might look like by illustrating how issues with the Prentice Criterion are relevant to EBM+'s guidelines on mechanistic evidence.

Explaining Why: A multidisciplinary approach to token causality

Phyllis Illari University College London

This paper aims to draw philosophical insights about causality from a collaborative project with Philosophy, Psychology and Computer Science: Explaining Why. We will explore the key methodological commitment of the project, which is to study causal reasoning using realistic vignettes which allow people to show ordinary causal and counterfactual reasoning in more familiar, natural settings.

Within philosophy, work on causality has often been driven by a small number of artificial 'toy examples' like two assassins shooting simultaneously, while computer science has been dominated by a few powerful modelling techniques, such as DAGs. Finally, the psychological literature has largely relied on simple experimental paradigms like the 'Monte Carlo scenario' of selecting balls
from an urn, with probabilities artificially set. Such work has distracted from the messy, uncertain process of making real decisions using real data, and from focus on the individual case such as managing chronic illness like diabetes.

We will describe study using realistic vignettes and show what can come up, such as domain differences in outcomes. For example, people prefer gradual causes of health, but will accept abrupt causes of ill health, like injury, strongly indicating that they are using background knowledge of the kinds of causes found in particular domains to influence causal reasoning.

Gender, Intervention, and Causal Inference in Clinical Judgment

Liyuan Jiao
Department of Philosophy, Durham University

Intervention-based approaches to causal inference, such as the Rubin Causal Model (RCM), ground causal claims in well-defined treatments and counterfactual contrasts (Rubin 1974). In the health sciences, investigators frequently appeal to socially structured factors such as gender to explain systematic disparities in diagnosis, evaluation, and healthcare delivery. The difficulty is not that gender lacks causal significance, but that it does not readily correspond to a well-defined intervention in the sense required by the RCM. Drug administration can be modeled as a discrete treatment, whereas assigning the same individual a different gender does not yield a comparably well-defined treatment contrast. This tension generates a question for causal inference: how can causal claims about gender be formulated and justified within an intervention-based framework?

This difficulty has motivated influential arguments, most notably by Holland (1986), that gender cannot function as a legitimate treatment variable within the RCM. At the same time, persistent gender-based disparities in medicine raise questions about how such effects can be causally inferred. A prominent empirical strategy responds by intervening not on gender itself, but on gender cues such as names, pronouns, and other perceptual markers under controlled conditions (Goldin and Rouse 2000; Moss-Racusin et al. 2012). In clinical studies using standardized or virtual patient descriptions, patients presenting identical symptoms receive systematically different evaluations and treatment recommendations depending on perceived gender cues (Stutts et al., 2010). By holding symptom presentation fixed while varying perceived gender, such studies abstract from biological differences and isolate the role of gendered interpretation in diagnostic reasoning.

The paper argues for three claims. First, in many diagnostic contexts, gender is causally operative as a socially structured category through a network of mediated mechanisms, one of which is perceived gender in diagnostic categorization. Clinical judgment often proceeds through processes of categorization and stereotype-mediated evaluation, in which patients classified as women may be regarded as more emotional or less credible reporters of pain. Second, cue-based interventions manipulate this causally operative node, namely perceived gender, thereby identifying a path-specific causal effect within a broader network of gender-related mechanisms. Such interventions do not replace gender with perceived gender; rather, they identify a mechanism through which gender exerts causal influence in diagnostic reasoning. Third, extrapolation from cue experiments to claims about gender's causal relevance is justified only if the experimentally isolated categorization mechanism also operates in ordinary clinical practice and is not systematically counteracted.

Under these conditions, cue-based evidence supports qualitative claims that gender is causally relevant to diagnostic evaluation, without requiring that gender itself be represented as a directly manipulable treatment within the RCM. The analysis clarifies the scope and limits of cue-based evidence and shows how intervention-based frameworks can support causal inference about socially structured factors that resist direct manipulation by identifying the mechanisms through which they operate.

References

Goldin, C. and Rouse, C. 2000. 'Orchestrating impartiality: The impact of "blind" auditions on female musicians', American Economic Review, 90(4), pp. 715–741.

Holland, P.W. 1986. 'Statistics and causal inference', Journal of the American Statistical Association, 81(396), pp. 945–960.

Moss-Racusin, C.A., Dovidio, J.F., Brescoll, V.L., Graham, M.J. and Handelsman, J. 2012. 'Science faculty's subtle gender biases favor male students', Proceedings of the National Academy of Sciences, 109(41), pp. 16474–16479.

Rubin, D.B. 1974. 'Estimating causal effects of treatments in randomized and nonrandomized studies', Journal of Educational Psychology, 66(5), pp. 688–701.

Stutts, L.A., Hirsh, A.T., George, S.Z. and Robinson, M.E. 2010. 'Investigating patient characteristics on pain assessment using virtual human technology', European Journal of Pain, 14(10), pp. 1040–1045.

Breastfeeding Promotion and the Limits of Evidence-Based Policy: Lessons for Evidential Pluralism

María Jiménez-Buedo
UNED
mjbuedo@fsof.uned.es

Saúl Pérez-González
University of Valencia
saul.perez@uv.es

Over the past fifty years, breastfeeding has been widely promoted by health authorities as a practice associated with numerous benefits for both infants and mothers. Public health campaigns across the globe have emphasized the idea that "breast is best," citing research suggesting that breastfeeding reduces the risk of several conditions, including Sudden Infant Death Syndrome (SIDS), allergies, asthma, eczema, obesity, and type II diabetes. Breastfeeding has also been associated with improved cognitive outcomes, including higher IQ scores and lower rates of attention deficit hyperactivity disorder (ADHD). These claims appeared to be supported by an extensive body of observational research in which breastfed children consistently showed better health and developmental outcomes than formula-fed children. Importantly, many of these studies attempted to control for parental socioeconomic status and other background characteristics that might confound the relationship between infant feeding practices and child outcomes.

Because breastfeeding practices cannot ethically be randomized at the individual level, experimental evidence is rare. In the 1990s, the Promotion of Breastfeeding Intervention Trial (PROBIT) (Kramer et al., 2001), led by the epidemiologist Michael Kramer in Belarus with support from the World Health Organization, approximated an experimental design by randomizing hospitals to implement a breastfeeding promotion program or to continue standard practices. The intervention increased breastfeeding rates in the treatment group. Early PROBIT results were often interpreted as experimental confirmation of breastfeeding's benefits, reporting modest reductions in gastrointestinal infections and atopic eczema during the first year of life. Yet the effects were small and several outcomes previously reported in observational research were not confirmed.

From the 2010s onward, researchers increasingly reassessed the evidence linking breastfeeding to long-term health and cognitive advantages. A central concern was residual selection bias. In many high-income societies, mothers who breastfeed tend to have higher education and greater resources than those who do not. Studies using sibling comparisons and richer measures of family environments suggest that many previously reported advantages—particularly cognitive and long-term health effects—are substantially reduced once family-level confounding factors are addressed. This evolving empirical landscape raises questions about the relationship between scientific evidence

and public policy. In this paper, we analyse the promotion of breastfeeding through the lens of recent philosophical discussions on evidence-based policy and evidential pluralism.

Evidential pluralism emphasizes integrating evidence of correlations and evidence of mechanisms to establish causal claims in medicine (Parkkinen et al., 2018). In the breastfeeding case, observational correlational evidence faces persistent problems of confounding and selection bias, while experimental evidence is difficult to obtain. Evidence about biological mechanisms—for example pathways through which breast milk might influence immune development—could complement statistical studies by helping to distinguish genuine causal effects from spurious correlations. However, implementing such an evidential pluralist strategy would face challenges. Evidence about the relevant mechanisms has often been limited or inconclusive. Moreover, identified mechanisms may operate heterogeneously across individuals depending on factors such as maternal physiology, breastfeeding duration, and caregiving practices. The breastfeeding case therefore illustrates both the potential value of evidential pluralism for evidence-based policy and the practical difficulties of applying it under conditions of limited and heterogeneous evidence.

References

Kramer, M. S., Chalmers, B., Hodnett, E. D., Sevkovskaya, Z., Dzikovich, I., Shapiro, S., Collet, J.-P., Vanilovich, I., Mezen, I., Ducruet, T., Shishko, G., Zubovich, V., Mknuik, D., Gluchanina, E., Dombrovskiy, V., Ustinovitch, A., Kot, T., Bogdanovich, N., Ovchinikova, L., … for the PROBIT Study Group. (2001). Promotion of Breastfeeding Intervention Trial (PROBIT)A Randomized Trial in the Republic of Belarus. JAMA, 285(4), 413–420. https://doi.org/10.1001/jama.285.4.413

Parkkinen, V.-P., Wallmann, C., Wilde, M., Clarke, B., Illari, P., Kelly, M. P., Norell, C., Russo, F., Shaw, B., & Williamson, J. (2018). Evaluating evidence of mechanisms in medicine: Principles and procedures. Springer.

Variables are not real, but causes in health are still useful

Samantha Kleinberg
Computer Science Department, Stevens Institute of Technology

Scientists make many causal claims about health, such as egg consumption increasing cholesterol or ultraprocessed foods causing higher energy intake. While most analysis focuses on whether a given causal relationship (such as between a dietary exposure and outcome) is well-supported, there is an underappreciated choice that shapes the validity and generalizability of causal inferences: how we define and represent cause and effect variables. Variables are not objective features of the world, but rather choices that can ultimately change causal conclusions.

A variable can be represented with varying focus (e.g, macronutrients, processing-level, or food names for diet) or specificity (e.g., discretizing continuous-valued glucose into ranges). Further, some relationships are only true in certain contexts (e.g. location), or vary dynamically due to other properties (e.g, age, BMI). Yet it is not clear whether such factors should be viewed as causal themselves or as background conditions, and there are practical challenges when combining these often static variables with longitudinal data to infer causal relationships.

Despite the difficulties of transforming data into variables, such choices influence both the correctness of causal models and how useful they are. Many works test narrowly pre-defined hypotheses, but this limits the generalizability of findings. If we only test for the effect of one specific food on a health outcome, we cannot generalize to whether other foods with similar nutrient composition or physical properties would have similar effects. Further, computational methods for finding causes from observational data assume variables are correctly specified, yet there is no guidance on what 'correct' means or how to achieve this.

In this paper I show how the practical realities of finding causes from data and using them for everyday health decisions raise philosophical questions about the nature of causality. I argue that there is no metaphysically "correct" set of variables, rather we impose constraints to provide order to our data. As a result, causal claims must be viewed as relative to these choices. Despite this, I show how pragmatic criteria can help determine which representations are better suited to a problem and argue relativity to a variable representation does not mean causal claims are arbitrary. Through case studies in nutrition, I show how computational tools can automatically identify appropriate granularity for variables and learn causal relationships that generalize across contexts. Studies with human subjects show that aligning variable representation with human needs for decision-making provides a more practical foundation for variable selection than interventionist approaches.

From Population- to Individual-Level Evidence: Two Concepts, Two Problems

Riin Kõiv
University of Barcelona
riin.koiv@ub.edu

Genetics research has been accumulating evidence about the genetic correlates of medical traits. There are increasingly calls and attempts to make use of this information (e.g., polygenic scores) a standard part of clinical practice. Such applications, however, continue to encounter criticisms. One common criticism goes as follows: existing empirical findings may provide evidence that a genetic variable is a population-level cause of a medical trait, however, such population-level evidence does not pass as evidence of the causes of individual instantiations of the trait — and it is the latter that is needed for preventing, predicting, diagnosing, and treating medical conditions in individual patients [1, 2, 3, 4]

This talk aims to clarify the nature of the inferential gap between population- and individual-level causal evidence. I argue that there are at least two different types of gaps involved, which are not always clearly distinguished from one another.

I begin by pointing out that "X (e.g., a genetic variable) is a population-level cause of Y (e.g., a medical trait)" often bundles together at least two intensionally and extensionally distinct propositions:

- X has a non-zero average causal effect on Y in some population P (Population Average)
- X is a population-specific cause of Y — that is, X has a given average effect on Y given the distribution of background conditions in some specific population P (Population Specific)

These two senses of "population-level cause" generate distinct inferential problems for drawing conclusions about the causes of a particular individual i having a given Y value.

The problem with evidence for Population Average: uncertainty

I first explain why, if we know our target individual i is a member of P, then information I such that I is evidence for Population Average is typically also evidence regarding whether X is a cause of i's Y value (assuming a Bayesian account of evidential relevance). However, typically it is evidence of little practical value, because it leaves us in a state of substantial uncertainty about whether X is a cause of i's Y value. This renders I of limited, although potentially valuable, medical use.

The problem with evidence for Population Specific: undefined reference class

That I is evidence for Population Specific raises an additional, more fundamental, obstacle to using I as evidence about the causes of i's Y value. A minimal condition

for evidence for Population Specific to also count as evidence regarding the cause of i's Y value is that we know i to be a member of P that Population Specific is indexed to. However, in most cases this is difficult to establish. The issue is not merely a lack of information about i. It is conceptual: in GWAS research, population membership is typically defined loosely in terms of proxy characteristics that only loosely track the causally relevant background conditions that Population Specific is indexed to. That is, most instances of Population Specific leave the relevant reference class un- or at least under-defined.

References

[1] Lynch, Kate E. 2021. "The Meaning of 'Cause' in Genetics." Cold Spring Harbor Perspectives in Medicine a040519. doi:10.1101/cshperspect.a040519.

[2] Madole, James W., and K. Paige Harden. 2023. "Building Causal Knowledge in Behavior Genetics." Behavioral and Brain Sciences 46:e182. doi:10.1017/S0140525X22000681.

[3] Serpico, Davide, and Mariusz Maziarz. 2023. "Averaged versus Individualized: Pragmatic N-of-1 Design as a Method to Investigate Individual Treatment Response." European Journal for Philosophy of Science 13(4):59. doi:10.1007/s13194-023-00559-0.

[4] Tabery, James. 2023. Tyranny of the Gene: Personalized Medicine and Its Threat to Public Health. Knopf.

Conflicting causal paths and randomized experiments

Mariusz Maziarz
Institute of Philosophy, Jagiellonian University
mariusz.maziarz@uj.edu.pl

Evidential Pluralism, a philosophical position that supports the use of both difference-making and mechanistic evidence, has developed tools to evaluate the quality of evidence for mechanisms understood as complex systems in medicine (Parkkinen et al., 2018). Some efforts have been made to implement the EP program in social sciences (Shan & Williamson, 2023; Maziarz, 2021). However, the position of EP has been criticized by supporters of Evidence-Based Medicine, an alternative that emphasizes evidence of difference-making from clinical trials. For example, Howick (2011) used some case studies of effective pharmaceutical treatments and non-pharmaceutical interventions that were successfully used in clinical practice despite their mechanisms of action remaining unknown at the time of their introduction. To defend EP as descriptively adequate for establishing causality in medicine, Williamson (2019) defended the EBM+ approach against accusations that some causal claims in medicine have been accepted based on associational studies only by claiming that those studies, by delivering difference-making evidence, also indicate that there is a mechanism linking a cause with its putative effect (see Stegenga 2022 for criticism). This point of view contradicts Illar's (2011) analysis, which holds that mechanistic evidence provides evidence for a specific mechanism (Mn) by indicating which entities, actions, and interactions constitute Mn (see Machamer, Darden & Craver2000).

This poses the question of whether experiments can provide mechanistic evidence for behavioral policies. Favereau and Nagatsu (2020a; 2020b; 2025) used examples of development policy experiments to argue that economists use lab-in-the-field experiments to discover causal mechanisms. I respond and argue, using examples, that randomized field experiments in economics, which instantiate black-box experiments, can provide evidence for specific mechanisms in certain contexts when background knowledge suggests a few possible mechanisms. Considering the purported similarity between medical RCTs and randomized field experiments (see Favereau 2016), I also discuss some examples of clinical trials designed to discriminate between two mechanistic hypotheses.

Additionally, I side with Illari (2011) on the grounds that randomized experiments may systematically report false-positive results when plagued by questionable research practices. On this basis, I argue that 'providing mechanistic evidence' should be narrowly understood as providing evidence for a specific mechanism, not for any possible mechanism linking a cause to its effect.

References

Favereau. (2016). On the analogy between field experiments in economics and clinical trials in medicine.

Favereau&Nagatsu. (2020b). Holding back from theory: limits and methodological alternatives of randomized field experiments in development economics.

Favereau&Nagatsu. (2025). Field experiments in economics: history and methodology.

Howick. (2011). Exposing the vanities—and a qualified defense—of mechanistic reasoning in health care decision making.

Illari. (2011). Mechanistic evidence: disambiguating the Russo–Williamson thesis.

Machamer et al. (2000). Thinking about mechanisms.

Maziarz. (2021). Resolving empirical controversies with mechanistic evidence.

Favereau&Nagatsu. (2020a). Two strands of field experiments in economics: A historical-methodological analysis.

Parkkinen et al. (2018). Evaluating evidence of mechanisms in medicine: principles and procedures.

Shan&Williamson. (2023). Evidential pluralism in the social sciences.

Stegenga. (2022). Evidence of effectiveness.

Williamson. (2019). Establishing causal claims in medicine.

Nothing but Causal Veneer: A Critique of Gut Microbiome Causation in Anorexia Nervosa

Jacob P. Neal
Department of Philosophy, University of Oregon
jpneal@oregon.edu

Nicolae Morar
Environmental Studies and Department of Philosophy, University of Oregon
nmorar@uoregon.edu

Recent research into the gut microbiome (GM) has claimed that specific microbial features are causally implicated in the development and maintenance of mental disorders. This paper critically evaluates these claims within the context of anorexia nervosa (AN), analyzing the influential study by Fan and colleagues (2023). While such research is often framed as a breakthrough in identifying new sites for therapeutic intervention, we argue that the claims of GM causation are largely unwarranted. Our analysis identifies the methodological and rhetorical strategies used by the authors to generate this hype by giving a causal veneer to data that remain fundamentally correlational.

The critique follows the two types of studies conducted by Fan and colleagues (2023). First, we analyze their use of mediation analysis to link GM features and metabolites to psychological constructs, such as "drive for thinness." We argue that the apparent validity of this analysis rests on the authors cherry-picking specific GM-metabolite correlations for which plausible mechanistic stories already exist in the literature. By highlighting a few correlations discovered via mediation analysis that appear plausibly causal, the authors create a veneer of causation for the remainder of associations, masking the likelihood that many are potentially spurious.

Within these associational studies, we further identify a shortcoming in the mechanistic stories provided. Even the most plausible causal pathways put forward by the authors, such as those that purport to link GM features to AN psychological constructs via metabolites affecting satiety, in fact, only connect GM features to physical observables like BMI or eating behavior. That is, these mechanisms offer no support for a causal link between the GM features identified and the complex psychological constructs themselves (e.g., "drive for thinness").

Second, we examine the fecal mouse transplant (FMT) interventions often cited as the gold standard for evidence of GM causation (Lynch et al., 2019). In the case of Fan and colleagues (2023), we argue that these experiments suffer from two major shortcomings. First, the study fails to track which microbes are successfully grafted into the mouse gut, and second, the scientists provide no animal model proxies for human mental states and instead track only BMI. Consequently, although the FMT does validate a connection

between GM features and weight regulation, it does not substantiate the stronger claim that the GM features cause the psychological hallmarks of AN.

Ultimately, we conclude that much of the purportedly causal evidence in gut microbiome studies of this sort is primarily hype (see also, Bourrat 2018). While GM research may serve as a useful heuristic in the context of discovery, framing it as discovering causes of AN is unwarranted.

References

Bourrat, P. 2018. "Have causal claims about the gut microbiome been over-hyped?", BioEssays, 40:e1800178

Fan, Yong, René Klinkby Støving, Samar Berreira Ibraim, et al. 2023. "The Gut Microbiota Contributes to the Pathogenesis of Anorexia Nervosa in Humans and Mice." Nature Microbiology 8 (5): 787–802.

Lynch, Kate E., Emily C. Parke, and Maureen A. O'Malley. 2019. "How Causal Are Microbiomes? A Comparison with the Helicobacter Pylori Explanation of Ulcers." Biology & Philosophy 34 (6): 62.

Rethinking Evidence in Healthcare: The Epistemic Value of Patient Testimony

Noemi Paciscopi
Università Vita-Salute San Raffaele

What should count as evidence in healthcare? This paper addresses this question by examining ongoing philosophical debates about the nature, scope, and normative structure of evidence in medicine. While evidence-based medicine (EBM) has significantly strengthened clinical research standards, its hierarchical model – typically privileging randomised controlled trials and statistical difference-making – rests on contestable philosophical assumptions about what qualifies as epistemically relevant support.

Recent discussions in the philosophy of science emphasise that no single type of evidence is sufficient to establish medical claims (Goldenberg, 2006; Russo and Williamson, 2007; Canali, 2019; Campaner and Galavotti 2012). Statistical correlations, mechanistic reasoning, clinical expertise, case-based knowledge, and patient testimony can contribute to justified belief. However, one form of evidence remains systematically under-theorised and often marginalised in practice: patients' first-person reports.

Drawing on accounts of evidence according to which something counts as evidence if it tends to make belief in a proposition reasonable or justified (Kim, 1988; Conee and Feldman 2004; Kelly 2006), I argue that first-person reports of pain, bodily change, treatment effects, and prior clinical experiences frequently satisfy evidential criteria in medical contexts. Their epistemic status should not be dismissed solely on the grounds of subjectivity. The routine downgrading of patient testimonies suggests that evidential hierarchies in healthcare are shaped not only by methodological considerations but also by background assumptions regarding credibility and authority.

Feminist social epistemology provides tools to analyse these assumptions. Work on epistemic injustice (Fricker 2007, 2017; Dotson 2011, 2012; Medina 2021; Carel and Kidd 2014, 2017; Freeman and Stewart, 2018) demonstrates how patients – particularly women and members of marginalised groups – are frequently undermined in their role as knowers. When testimonial contributions are trivialised or pre-emptively discredited, the evidential base of clinical decision-making is narrowed in ways that may reduce epistemic reliability.

Drawing on qualitative empirical research in oncology and obstetrics (blind, blind), this study illustrates how evidential standards operate in practice and how patients' experiential reports are often treated as secondary to formally established research procedures. These patterns suggest that the problem is not merely one of insufficient data, but of a restricted conception of what counts as legitimate evidence.

This work argues for a reconstruction of EBM that recognises the epistemic value of multiple forms of evidence, including patient testimony. Broadening the evidential landscape in this way is not a departure from scientific rigour, but a refinement of it. By critically examining the normative assumptions underlying hierarchical evidence

structures, we can develop a more adequate account of medical evidence – one that is both epistemically more reliable and ethically more inclusive.

References

Canali, S. (2019). Evaluating evidential pluralism in epidemiology: Mechanistic evidence in exposome research. History and Philosophy of the Life Sciences, 41(1), 4.

Campaner, R., & Galavotti, M. C. (2012). Evidence and the assessment of causal relations in the health sciences. International Studies in the Philosophy of Science, 26(1), 27–45.

Carel, H., & Kidd, I. J. (2014). Epistemic injustice in healthcare: a philosophical analysis. Medicine, Health Care and Philosophy, 17(4), 529–540.

Carel, H., & Kidd, I. J. (2017). Epistemic injustice in medicine and healthcare. In The Routledge handbook of epistemic injustice (pp. 336–346). Routledge. Conee, E., & Feldman, R. (2004). Evidentialism: Essays in epistemology. Clarendon Press.

Dotson, K. (2011). Tracking epistemic violence, tracking practices of silencing. Hypatia, 26(2), 236–257.

Dotson, K. (2012). A cautionary tale: On limiting epistemic oppression. Frontiers: A Journal of Women Studies, 33(1), 24–47.

Freeman, L., & Stewart, H. (2018). Microaggressions in clinical medicine. Kennedy Institute of Ethics Journal, 28(4), 411–449.

Fricker, M. (2007). Epistemic injustice: Power and the ethics of knowing. Clarendon Press.

Fricker, M. (2017). Evolving concepts of epistemic injustice. In The Routledge handbook of epistemic injustice (pp. 53–60). Routledge.

Goldenberg, M. J. (2006). On evidence and evidence-based medicine: lessons from the philosophy of science. Social Science & Medicine, 62(11), 2621–2632.

Kelly, T. (2016). Evidence. In E. N. Zalta (Ed.), The Stanford Encyclopedia of Philosophy, winter 2016 edn. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/entries/evidence/

Kim, J. (1988). What is 'naturalized epistemology?' Philosophical Perspectives, 2, 381–405. https://doi.org/10.2307/2214082

Medina, J. (2021). Agential epistemic injustice and collective epistemic resistance in the criminal justice system. Social Epistemology, 35(2), 185–196.

Russo, F., & Williamson, J. (2007). Interpreting causality in the health sciences. International Studies in the Philosophy of Science, 21(2), 157–170.

Mechanistic discarding and target-to-target extrapolation

Guillermo Marín Penella
University of the Balearic Islands

Losner Briones
UNED

The extrapolation of causal claims plays a central role in regulatory and policy decision-making, particularly under conditions of epistemic constraints and risk. In health regulatory science, establishing the efficacy and safety of pharmaceutical interventions often faces significant obstacles, since ethical and legal restrictions may preclude direct testing in human populations. To address these limitations, regulatory agencies routinely integrate extrapolative reasoning, typically mobilising data from source-as-model populations (A) to inform causal claims about target-of-similarity populations (B). A paradigmatic case arises in drug development, where evidence from non-human animal models is extrapolated to initial human cohorts (ICH, 2009; Lemoine, 2017).

Within this domain, Evidential Pluralism holds that evidence of mechanisms can justify similarity between populations by helping to bridge the inferential gap created by the inaccessibility of population B and the corresponding lack of correlational evidence in B (Parkkinen et al., 2018; Shan and Williamson, 2023). Thus far, the pluralistic account of causal extrapolation has identified four strategies for justifying mechanistic similarity and two principal contexts of application (Parkkinen and Willianson, 2020; Williamson, 2019). These strategies—enumerative induction, robustness analysis, phylogenetic reasoning, and comparative process tracing—primarily aim to establish shared mechanisms, while the recognised contexts are confined to two stages of pharmaceutical development: animal-to-human and study-to-target populations. This paper expands the pluralistic framework along both dimensions.

First, we show that regulatory practice also relies on mechanistic evidence— particularly evidence of interactions between mechanisms—to rule out causal differences between population B and population A. We term this new strategy 'mechanistic discarding.' Our central example is the European Medicines Agency's (EMA) safety extrapolation of the COVID-19 vaccine Vaxzevria for pregnant individuals, a population excluded from pivotal RCTs. The EMA (n.d.; 2021) judged that maternal immunosuppression would not interfere with the causal pathway linking vaccination to safety outcomes. Given mechanistic knowledge of Vaxzevria as a non-replicating viral-vector vaccine, the agency considered the interaction with a suppressed immune system unlikely to generate adverse effects (EMA, 2022).

This case supports a broader implication: pluralistic causal extrapolation may proceed via a negative route, namely through the eliminative induction of potentially relevant causal differences bearing on cross-population heterogeneity, together with inference to the best explanation of their causal irrelevance grounded in mechanistic similarity.

Second, we argue that pluralistic causal extrapolation also operates within phase IV post-marketing surveillance, where both the source-as-model and target-of-treatment populations are real-world therapeutic populations. We term this new context 'target-to-target extrapolation' and argue that it differs from the previously recognised contexts in two fundamental respects. First, in relation to animal-to-human extrapolations, it shifts the focus towards real-world effectiveness and safety, rather than concentrating on toxicity thresholds or dose optimisation. Second, concerning study-to-target extrapolations, it usually rests on a hybrid evidential base integrating results from experimental sources with real-world data drawn from observational infrastructures such as electronic health records or clinical registries.

We illustrate target-to-target extrapolation through the U.S. Food and Drug Administration's authorisation of Palbociclib for the treatment of male breast cancer via extrapolation of its safety evidence from female populations (CDER, 2017, 2019; Wedam et al., 2020). The case shows that Evidential Pluralism can be applied beyond strictly experimental settings and centrally involve the integration of real-world evidence.

References

CDER. (2017). Approval Package for: APPLICATION NUMBER 207103Orig1s004. Center for Drug Evaluation and Research. Retrieved March 8, 2025, from https://www.accessdata.fda.gov/drugsatfda_docs/nda/2019/207103Orig1s004.pdf

CDER. (2019). Approval Package for: APPLICATION NUMBER 207103Orig1s008. Center for Drug Evaluation and Research. Retrieved March 5, 2025, from https://www.accessdata.fda.gov/drugsatfda_docs/nda/2019/207103Orig1s008.pdf

EMA. (2021). Assessment report. COVID-19 Vaccine AstraZeneca. Common name: COVID-19 Vaccine (ChAdOx1-S [recombinant]). Procedure No. EMEA/H/C/005675/0000. European Medicines Agency. Retrieved November 17, 2025, from https://www.ema.europa.eu/en/documents/assessmentreport/vaxzevria-previously-covid-19-vaccine-astrazeneca-epar-public%20assessment-report_en.pdf

EMA. (2022). Vaxcevria (COVID-19 Vaccine (ChAdOx1-S [recombinant])). European Medicines Agency. Retrieved February 4, 2025, from https://www.ema.europa.eu/en/documents/overview/vaxzevria-previously-covid%2019-vaccine-astrazeneca-epar-medicine-overview_en.pdf

EMA. (n.d.). Vaxzevria (previously COVID-19 Vaccine AstraZeneca). European Medicines Agency. Retrieved November 2024, from https://www.ema.europa.eu/en/medicines/human/EPAR/vaxzevria

ICH. (2009). Guidance on nonclinical safety studies for the conduct of human clinical trials and marketing authorization for pharmaceuticals M3(R2). International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use.

Retrieved December 17, 2022, from https://database.ich.org/sites/default/files/M3_R2__Guideline.pdf

Lemoine, M. (2017). Animal extrapolation in preclinical studies: An analysis of the tragic case of TGN1412. Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, 61, 35–45. https://doi.org/10.1016/j.shpsc.2016.12.004

Parkkinen, V.-P., Wallmann, C., Wilde, M., Clarke, B., Illari, P., Kelly, M. P., Norell, C., Russo, F., Shaw, B., and Williamson, J. (2018). Evaluating Evidence of Mechanisms in Medicine: Principles and Procedures. Springer International Publishing. https://doi.org/10.1007/978-3-319-94610-8

Parkkinen, V.-P. and Williamson, J. (2020). Extrapolating from Model Organisms in Pharmacology. In A. LaCaze & B. Osimani (Eds.), Uncertainty in Pharmacology (pp. 59–78). Springer International Publishing. https://doi.org/10.1007/978-3-030-29179-2

Shan, Y. and Williamson, J. (2023). Evidential Pluralism in the Social Sciences. Routledge. https://doi.org/10.4324/9781003143000

Wedam, S., Fashoyin-Aje, L., Bloomquist, E., Tang, S., Sridhara, R., Goldberg, K. B., Theoret, M. R., Amiri-Kordestani, L., Pazdur, R., and Beaver, J. A. (2020). FDA Approval Summary: Palbociclib for Male Patients with Metastatic Breast Cancer. Clinical Cancer Research, 26(6), 1208–1212. https://doi.org/10.1158/10780432.CCR-19-2580

Williamson, J. (2019). Establishing causal claims in medicine. International Studies in the Philosophy of Science, 32(1), 33–61. https://doi.org/10.1080/02698595.2019.1630927

Economic inequality and health outcomes: spelling out the causal pathways

Elena Popa
University of Seville

Jay Zameska
University of Southern Denmark

It is relatively uncontroversial *that* economic inequality has negative effects on health (Pickett & Wilkinson 2015, Wilkinson & Pickett 2009), yet less has been said regarding *how* the two relate. Recent work on population health has highlighted the need to act on the so-called fundamental causes (Link & Phelan 1995, Valles 2018), many of which are of an economic nature, but thus far there is no straightforward conceptualization of causality for these kinds of claims. We argue that spelling out the causal story is not just a technical concern but is also a normative one (cf. Popa & Zameska 2025). The way we conceptualize the causal link between inequality and health has direct implications for justice and priority setting, especially in a global context. Taking a step towards building such an account, in this paper we look at two ways of conceptualizing the effects of inequality, connecting them to competing theories of justice (Voigt and Wester 2015) and their implications for action.

Difference-making approaches to causation, such as interventionist or counterfactual ones, that have been popular in many domains can link patterns of the distribution of economic resources and health outcomes. This can provide insights for redistribution policies as economic interventions that also affect health. This approach aligns with a distributive conception of justice, and conceives of injustice primarily as a problem of maldistribution: economic inequalities are unjust because they deny individuals a fair share of the goods necessary to achieve health. It provides normative and evidential support for policies of economic redistribution by framing them as health interventions.

Secondly, one may choose to look more in-depth at social patterns obtaining in more economically unequal societies: relations characterized by patterns of domination and distrust, affective injustice (Gallegos 2021, Thomas et al. 2024), social capital (Szreter & Woolcock 2004). A more qualitative approach, using mechanistic or dispositional notions of causality, is better suited for singling out the precise pathways of these social patterns. This aligns with relational conceptions of justice, which similarly locate injustice in these patterns of social relations rather than in resource distribution. From this perspective, the primary injustice is not a distributive problem, but a social hierarchy that leads to diminished health. This view provides normative and evidential support for policies that aim to encourage egalitarian social relations, such as promoting particular forms of interaction (e.g., fostering social support or facilitating access to public services).

This paper shows how different causal models of health inequality support different normative arguments for policy action. Causal claims based on resource distribution provide evidence for interventions justified by distributive justice. In contrast, causal

claims based on social hierarchies provide evidence for interventions justified by relational justice. Distinguishing between these causal and normative arguments is a necessary step to clarify policy debates on health equity. This approach demonstrates that a precise account of causation is necessary for a precise account of justice, which can then inform priority setting (Goldberg 2022).

References

Gallegos, F. (2021). Affective injustice and fundamental affective goods.

Goldberg, D. (2022). Fundamental Cause Theory. In The Routledge Handbook of Philosophy of Public Health. Taylor & Francis, pp. 85-97.

Link, B. G., & Phelan, J. (1995). Social conditions as fundamental causes of disease. Journal of health and social behavior, 80-94.

Pickett, K. E., & Wilkinson, R. G. (2015). Income inequality and health: a causal review. Social science & medicine, 128, 316-326.

Popa, E., & Zameska, J. (2025). Causal pluralism and public health ethics trade-offs: a toolkit for acting on social determinants of health. Philosophy Compass, 20(11), e70064.

Szreter, S., & Woolcock, M. (2004). Health by association? Social capital, social theory, and the political economy of public health. International journal of epidemiology, 33(4), 650-667.

Thomas, A., Archer, A., & Engelen, B. (2024). Extravagance and misery: the emotional regime of market societies. Oxford University Press.

Valles, S. (2018). Philosophy of population health: Philosophy for a new public health era. Routledge.

Voigt, K., & Wester, G. (2015). Relational equality and health. Social Philosophy and Policy, 31(2), 204-229.

Wilkinson, R.G., and Pickett, K. (2009) The spirit level: Why more equal societies almost always do better. Vol. 6. London: Allen Lane.

Reframing causal modelling using observational data: from actionability to evidence

Stella Prizeman-Green
Usher Institute, University of Edinburgh

While statistical methods for gaining causal knowledge from observational data (referred to here as "causal models") have gained popularity in the healthcare setting, they are not yet mainstream, with many journals still restricting the use of causal language for observational studies [1]. This stance as an outsider methodology has precipitated an environment in which causal models must be "sold" to broader audiences in the clinical literature, typically via their proposed benefits over other popular methodologies. For example: causal models provide a practical alternative when a randomised trial is not possible (e.g. [2]), or causal models provide a more "actionable" alternative to predictive models (e.g. [3]). I argue that this latter point risks replicating mistakes made in the promotion of AI-based clinical decision support tools, obscuring genuine benefits of causal models as well as their limitations. I propose that framing causal models as a source of evidence better characterises their epistemic contribution and enables productive critical reflection on their limitations. This work is informed by my positionality as an applied data science researcher with a background in philosophy, and I employ examples from my own data science practice in the critical care setting to motivate my arguments.

The argument from actionability is typically posed as: predictive models forecast outcomes, but they do not provide any information on how that outcome might be changed. By contrast, since causal models explicitly compare counterfactual outcomes given an intervention, they provide actionable insight into how clinicians might change an outcome, thus helping them to make "better decisions" [4]. However, I make the case that given minimal assumptions about what constitutes a "better decision" in healthcare, it is unclear that causal models actually deliver this. Clinical decision-making is human, ethical and complex, requiring the reconciliation of different kinds of knowledge (e.g. patient preference, clinical experience, scientific evidence) as well as tools for reasoning under the uncertainty which is a daily reality in clinical practice [5]. Using examples from the clinical literature on causal methods (e.g. [6]) and my own data science practice, I argue that the scope of causal models' contribution to delivering "better decisions" is much more limited than the argument from actionability implies. However, this does not obviate their value: I contend that aiming for actionability risks replicating the various ethical and implementation challenges associated with the pursuit of autonomy in AI-based clinical decision support tools [7, 8].

Instead, I propose that framing causal models as a source of evidence better characterises both their benefits and limitations. When viewed as one potential source of evidence, to be synthesised with other forms of knowledge and methods for handling uncertainty, causal models' contribution to the clinical decision-making pathway is clarified.

This highlights the limits of their contribution to "better decisions" and re-emphasises clinicians' epistemological responsibility [9]. It also better situates causal models within broader frameworks of triangulation of evidence and causal pluralism, which I argue carries ethical and pragmatic benefits.

References

[1] Dahabreh IJ, Bibbins-Domingo K (2024) Causal Inference About the EYects of Interventions From Observational Studies in Medical Journals. JAMA 331:1845–1853. https://doi.org/10.1001/jama.2024.7741

[2] Hernán MA, Robins JM (2016) Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. American Journal of Epidemiology 183:758–764. https://doi.org/10.1093/aje/kwv254

[3] Prosperi M, Guo Y, Sperrin M, et al (2020) Causal inference and counterfactual prediction in machine learning for actionable healthcare. Nat Mach Intell 2:369–375. https://doi.org/10.1038/s42256-020-0197-y

[4] Hernán MA, Hsu J, Healy B (2019) A Second Chance to Get Causal Inference Right: A Classification of Data Science Tasks. CHANCE 32:42–49. https://doi.org/10.1080/09332480.2019.1579578

[5] Simpkin AL, Schwartzstein RM (2016) Tolerating Uncertainty — The Next Medical Revolution? New England Journal of Medicine 375:1713–1715. https://doi.org/10.1056/NEJMp1606402

[6] Maley JH, Wanis KN, Young JG, Celi LA (2020) Mortality prediction models, causal eYects, and end-of-life decision making in the intensive care unit. BMJ Health Care Inform 27:e100220. https://doi.org/10.1136/bmjhci-2020-100220

[7] Sokol K, Fackler J, Vogt JE (2025) Artificial intelligence should genuinely support clinical reasoning and decision making to bridge the translational gap. npj Digit Med 8:345. https://doi.org/10.1038/s41746-025-01725-9

[8] Van Cauwenberge D, Van Biesen W, Decruyenaere J, et al (2022) 'Many roads lead to Rome and the Artificial Intelligence only shows me one road' : an interview study on physician attitudes regarding the implementation of computerised clinical decision support systems. BMC MEDICAL ETHICS 23:. https://doi.org/10.1186/s12910-022-00787-8

[9] van Baalen S, Boon M (2015) An epistemological shift: from evidence-based medicine to epistemological responsibility. Journal of Evaluation in Clinical Practice 21:433–439. https://doi.org/10.1111/jep.12282

Circularity and reification in HiTOP: the problem of causal independence

Rosa Rooduijn
Rijksuniversiteit Groningen

Recently, there has been a proliferation of "data-driven" approaches in psychiatry. One of them is the *Hierarchical Taxonomy of Psychopathology* (HiTOP). HiTOP presents itself as a "purely empirical" alternative to the currently dominant taxonomic system of the *Diagnostic and Statistical Manual of Mental Disorder* (DSM). An often-voiced criticism to the DSM is that it fosters the reification of its diagnostic categories. One form of reification cited is the invocation of DSM diagnostic categories as causes of their symptoms. In this paper, I argue that the HiTOP framework repeats this form of causal reification. More broadly, the analysis highlights a recurring risk in data-driven medicine: the appearance of purely empirical discovery can obscure rather than resolve foundational problems of causal interpretation.

First, I set out the conceptual relation between reification, causal claims, and the requirement of independence between cause and effect. Drawing on interventionist and difference-making accounts of causation, I argue that invoking abstract constructs in causal claims is not inherently illegitimate. Problems arise, however, when there is no adequate independence between cause and effect. In the paper, I make a distinction between *conceptual* independence and *epistemic* independence. Without conceptual independence between cause and effect, causal claims are circular. Without epistemic independence, there is no circularity in a strict sense, but it does render causal claims uninformative. I argue that the interpretation of DSM's diagnostic categories as causally responsibly for their symptoms does not fulfil the requirement of conceptual independence between cause and effect. Since the diagnostic categories refer to operationalized sets of symptom-based criteria, causal interpretations are circular.

Then, I argue that the HiTOP framework runs into the same causal circularity. In HiTOP, higher-order latent factors are extracted from the covariation of psychopathology symptoms. Although HiTOP presents its higher-order factors as descriptive summaries of symptom covariation, the reliance on reflective factor models strongly invites a causal interpretation. Furthermore, I show how the HiTOP literature routinely embeds these factors in etiological and entity-like language, portraying them as drivers of illness course and potential treatment targets. I show, however, that the latent factors extracted from the symptom covariation in HiTOP are exhaustively constructed from the very symptoms they are supposed to cause. Hence, causal interpretations violate the condition of independence between cause and effect. Generally, the violation concerns conceptual independence, because the latent factors are not defined independently of the symptoms they are purported to cause. In some instances, attempts are made to attach some substantial meaning to the latent factors beyond the symptom covariation, but in those cases a violation of epistemic independence remains. These violations of causal independence are obscured by their embedding in statistical techniques and their presentation as empirical findings.

I conclude that HiTOP does not eliminate the theoretical commitments it aims to transcend but instead renders them less visible. Compared with the DSM, whose reification can be traced to a medical framework, HiTOP's reification is more insidious because it is mediated by ostensibly neutral statistical techniques.

Distinct Social Epistemological Structures of Causal Inference in Clinical and Basic Medical Research

Tatsuyuki Sato
Hokkaido University

Clinical medical research and basic biomedical research are two pillars of medical science. Clinical research typically investigates the efficacy of interventions in patients, while basic research investigates the mechanisms of those effects. Although this division is familiar in practice and has been articulated in discussions of evidential and methodological pluralism, there has been little investigation from the perspective of social epistemology. Here, I argue that the two research fields differ not merely in subject matter but also in the location of the core inferential step within the social epistemological structure.

Clinical research, most typically randomized controlled trials, aims to establish statistical associations between exposures and outcomes. A typical study reports the number of participants, the number of outcomes, and statistical measures such as effect sizes and p-values, but refrains from making general claims, such as "drug D is effective for condition C." The inferential step from data to generalized claims and clinical recommendations is intentionally deferred to collectives such as guideline committees, a structure I refer to as an *outsourcing of inference*. Individual researchers are responsible for producing high-quality statistical evidence, while general causal claims are proposed by collective bodies in the form of guidelines and statements.

Basic research investigates mechanisms of disease and physiological processes. It aims to understand mechanisms by integrating multiple experiments and constructing models of underlying biological processes, such as "molecule X phosphorylates molecule Y, thereby improving cardiac function." In basic research, individual researchers not only report the data as it is, but also propose these mechanistic models in individual papers. This means that the crucial inferential work involving model construction and causal interpretation occurs at the level of individual researchers. The community serves as a safeguard to assess whether the inferences made by individual researchers are plausible, but it does not itself introduce causal interpretations. Theoretical commitments are introduced at the individual level rather than the collective level.

In philosophy of science, there have been extensive discussions of evidential pluralism for causality, primarily focusing on the balance between statistical and mechanistic evidence. My account adds a novel layer of argument to the issue from the perspective of social epistemology, showing that the two types of evidence are not only different in their subject matter but also in the social structures in which they are produced.

This account has practical implications for efforts to improve scientific practice. Preregistration has become the norm in clinical research, and there is now a movement to extend it to basic biomedical research to address the reproducibility crisis. However,

these efforts have not advanced as much as expected. The argument proposed here may suggest that preregistration is well-suited to clinical research, where individual researchers can focus on reporting what was done and what was observed, but notsuitable for basic research, where data curation and inference proceed in an entangled manner within individual researchers. Improving medical science, therefore, may require moving beyond a one-size-fits-all approach and instead tailoring strategies depending on the social epistemological structures of the target type of research.

From population evidence to individual causal inference: evidential pluralism and the epistemology of N-of-1 medicine

Laura Senis Villacé
University of Valencia

Evidence-based medicine (EBM) has traditionally relied on causal inferences derived from randomized controlled trials (RCTs) and other population-level statistical studies. While this approach has been enormously successful for evaluating average treatment effects, it raises a persistent epistemological challenge: how can causal claims justified at the population level be legitimately applied to individual patients? This problem has become increasingly pressing with the rise of precision and N-of-1 medicine, where clinical decisions are guided by detailed individual data streams (e.g., wearable devices, continuous glucose monitoring, or digital phenotyping) and where treatment effects may vary substantially across patients (Schork, 2015; Natesan et al. 2023; Hawksworth et al. 2024).

This paper examines the philosophical implications of this shift for causal inference in the health sciences. I argue that the epistemology underlying traditional EBM—centered on difference-making evidence from RCTs—faces limitations when applied to individual-level clinical reasoning. In N-of-1 contexts, clinicians increasingly rely on heterogeneous forms of evidence, including longitudinal intra-individual data, mechanistic knowledge about disease pathways, and population-level probabilistic information. This situation raises the question of how these distinct evidential sources can jointly support causal claims about what works for a particular patient.

Building on the framework of evidential pluralism developed by Russo and Williamson (2007) and further elaborated in philosophy of medicine (Clarke et al. 2014), I propose that N-of-1 medicine provides a compelling case study for extending pluralistic accounts of causal evidence to the level of individual inference. In particular, I suggest that causal reasoning in such contexts integrates three epistemic components: (1) population-level probabilistic evidence about average treatment effects, (2) mechanistic evidence concerning biological pathways relevant to the individual patient, and (3) temporally structured intra-individual data that track changes before and after interventions. Rather than treating these evidential sources as hierarchically ordered—as in standard EBM hierarchies—they should be understood as complementary components of a pluralistic causal framework.

To illustrate this argument, I discuss emerging clinical practices involving continuous health monitoring technologies, such as wearable sensors and digital biomarkers, which enable repeated within-person observations and adaptive treatment adjustments. These practices effectively generate individualized experimental settings that blur the boundary between clinical care and experimentation.

The paper contributes to recent debates in philosophy of medicine concerning the limits of population-based inference (Cartwright 2012; Stegenga 2018) and the role of mechanisms in causal reasoning (Illari and Williamson 2012). By focusing on N-of-1 medicine, it highlights a domain where pluralistic approaches to causal evidence are not merely philosophically attractive but practically necessary. More broadly, the analysis suggests that contemporary developments in personalized and digital medicine may require rethinking the epistemic foundations of causal inference in the health sciences.

References

Cartwright, N. (2012). Evidence-Based Policy: A Practical Guide to Doing It Better. Oxford University Press.

Clarke, B., Gillies, D., Illari, P., Russo, F., & Williamson, J. (2014). Mechanisms and the Evidence Hierarchy. Oxford University Press.

Hawksworth, O., Chatters, R., Julious, S., et al. (2024). A methodological review of randomised n-of-1 trials. Trials, 25, 263.

Illari, P. M., & Williamson, J. (2012). What Is a Mechanism? Thinking about Mechanisms Across the Sciences. European Journal for Philosophy of Science, 2(1), 119–135.

Natesan Batley, P., McClure, E. B., Brewer, B., Contractor, A. A., Batley, N. J., Hedges, L. V., & Chin, S. (2023). Evidence and reporting standards in N-of-1 medical studies: a systematic review. Translational Psychiatry, 13, 263.

Russo, F., & Williamson, J. (2007). Interpreting causality in the health sciences. International Studies in the Philosophy of Science, 21(2), 157–170.

Samuel, J. P., Wootton, S. H., Holder, T., & Molony, D. (2022). A scoping review of randomized trials assessing the impact of n-of-1 trials on clinical outcomes.

Schork, N. J. (2015). Personalized medicine: Time for one-person trials. Nature, 520, 609–611.

Stegenga, J. (2018). Medical Nihilism. Oxford University Press.

Vogt, H., & Hofmann, B. (2022). How precision medicine changes medical epistemology: A formative case from Norway. Journal of Evaluation in Clinical Practice, 28(6), 1205–1212.

Williamson, J. (2019). Establishing causal claims in medicine. International Studies in the Philosophy of Science, 33(1), 33–61.

Causality and the Network Theory of Psychopathology

Anna van Oosterzee
Utrecht University
Leiden University
Ethics of Socially Disruptive Technologies (ESDiT) Consortium

Cognitive neuroscience has thus far not identified common causal mechanisms that account for the DSM's symptom profiles. In parallel, many contemporary data-driven models in psychiatry improve prediction and may support clinical control (e.g., forecasting relapse risk or informing treatment selection), yet they rarely deliver what the medical model demands from causal explanation: validation of an underlying disease entity. This motivates a central question for causal inference in psychiatry: if DSM-classified disorders cannot be substantiated as brain diseases with shared mechanisms, what would count as a causal understanding of psychopathology?

A prominent alternative rejects the assumption that people with the same diagnosis share one common cause or causal structure. Instead, mental disorders may be better understood as "complex, mutually reinforcing networks of causal mechanisms," i.e., higher-order disturbances in multi-level systems (Kendler, 2012). This idea has been most thoroughly developed in the network approach to psychopathology (Borsboom, 2008; Borsboom et al., 2019; Borsboom & Cramer, 2013; Fried & Cramer, 2017). On this view, disorders are not latent entities that cause symptoms; rather, disorders are constituted by patterns of spreading activation among symptoms themselves. Crucially, the network approach aims to provide causal explanations without appealing to a single underlying mechanism or common cause (Kalis & Borsboom, 2020).

In this talk I assesses whether network theory can be a genuine explanatory alternative to the medical common-cause model. I interpret Borsboom's proposal as a form of *topological explanation*: an explanation that abstracts from lower-level details and instead explains system behaviour by appeal to global connectivity patterns. However, whether topological models genuinely explain, rather than merely redescribe, remains contested. Mechanistic "ontic" approaches argue that network explanations require nodes and edges to correspond to the right kinds of causal components and relations (Craver, 2016; Craver & Povich, 2017), whereas autonomist accounts defend the possibility of explanatory topology without reduction to lower-level mechanisms, provided the model supports appropriate counterfactual reasoning (Kostic, 2025; Kostic & Khalifa, 2023).

I argue that the network approach faces two pressing challenges. First, *node selection*: DSM symptoms are pragmatically convenient but unlikely to be the correct causal building blocks, risking omission of relevant behavioural, social, physiological, and digital-phenotype variables. Second, *circularity*: if symptoms both define the disorder and serve as its explanans, the target of explanation must be redescribed as *dynamic behaviour* (onset, maintenance, tipping, relapse) rather than diagnostic category membership. I conclude that, when networks are developed to support contrastive, coun-

terfactual dependencies, exemplified by vulnerability hypotheses linking connectivity to tipping and hysteresis, topological properties can function as difference-makers. On a permissive mechanistic reading, symptom networks can therefore contribute genuine causal explanations in psychiatry without relying on a single common cause.

References

Borsboom, D. (2008). Psychometric perspectives on diagnostic systems. Journal of Clinical Psychology, 64(9), 1089–1108. https://doi.org/10.1002/jclp.20503

Borsboom, D., & Cramer, A. (2013). Network Analysis: An Integrative Approach to the Structure of Psychopathology. Annual Review of Clinical Psychology, 9, 91–121. https://doi.org/10.1146/annurev-clinpsy-050212-185608

Borsboom, D., Cramer, A. O. J., & Kalis, A. (2019). Brain disorders? Not really: Why network structures block reductionism in psychopathology research. Behavioral and Brain Sciences, 42, e2. https://doi.org/10.1017/S0140525X17002266

Craver, C. F. (2016). The Explanatory Power of Network Models. Philosophy of Science, 83(5), 698–709. https://doi.org/10.1086/687856

Craver, C. F., & Povich, M. (2017). The directionality of distinctively mathematical explanations. Studies in History and Philosophy of Science Part A, 63, 31–38. https://doi.org/10.1016/j.shpsa.2017.04.005

Fried, E. I., & Cramer, A. O. J. (2017). Moving Forward: Challenges and Directions for Psychopathological Network Theory and Methodology. Perspectives on Psychological Science, 12(6), 999–1020. https://doi.org/10.1177/1745691617705892

Kalis, A., & Borsboom, D. (2020). Folk psychology as a causal language. Theory & Psychology, 30(5), 723–728. https://doi.org/10.1177/0959354320933940

Kendler, K. S. (2012). Levels of explanation in psychiatric and substance use disorders: Implications for the development of an etiologically based nosology. Molecular Psychiatry, 17(1), 11–21. https://doi.org/10.1038/mp.2011.70

Kostic, D. (2025). Does functional connectivity explain? Manuscript Submitted for Publication.

Kostic, D., & Khalifa, K. (2023). Decoupling Topological Explanations From Mechanisms. Philosophy of Science, 90(2), 245–268. https://doi.org/10.1017/psa.2022.29

Causal Inference Beyond Evidence-Based Medicine: A Bayesian Appraisal of Mechanistic Plausibility

Ethan Vorster
IUSS Pavia

Despite longstanding criticisms, EBM remains the dominant framework for evaluating causal claims in healthcare. It privileges population-level statistical evidence from randomised controlled trials (RCTs) and meta-analyses while downgrading mechanistic evidence and pathophysiological reasoning. Although this hierarchy has proven effective in the context of pharmaceutical interventions, it struggles when applied to interventions whose therapeutic effects are delivered through complex engineered systems rather than biochemical agents. I call this class of interventions machine-based medicine (MBM). MBM includes, but is not limited to, extracorporeal membrane oxygenation, ventilation, dialysis, MRI-guided focused ultrasound, X-ray therapy and particle-beam therapy (PBT). What unifies these technologies is a distinctive evidential profile: typically, they are supported by a strong biophysical rationale and robust mechanistic evidence grounded in physics, engineering, and physiology, yet they often lack the high-level RCT evidence privileged by EBM. This deficit is not due to the epistemic irrelevance of RCTs but to ethical, logistical, and financial constraints that make large-scale, high-quality RCTs difficult or impractical. As a result, MBM interventions often have high causal plausibility with constrained statistical evidence—an evidential profile that fits poorly within EBM's hierarchy. Using PBT as a paradigmatic case study within MBM, I examine this tension in detail. PBT exploits the Bragg Peak to deliver highly conformal doses of radiation to tumours while sparing surrounding healthy tissue. The biophysical rationale is straightforward: reduced irradiation of healthy tissue should lower the risk of long-term complications and improve quality of life compared to conventional X-ray therapy. This rationale is supported by radiobiological studies, dosimetric-biological correlations, imaging data, biomarker research, and probabilistic modelling, forming a compelling—though heterogeneous—mechanistic picture. Despite this strong causal plausibility, the statistical evidence for PBT remains uneven. For many rare and paediatric cancers, outcomes show clear reductions in adverse effects, whereas evidence for common adult cancers is often inconclusive. High-quality RCTs remain scarce due to problems of equipoise, patient preferences, long follow-up periods, technological obsolescence, geographical scarcity of treatment centres, and the difficulty of designing trials that test for lack of adverse effects. Consequently, PBT is often described as "lacking evidence," however, a more accurate characterisation is that it exhibits a heterogeneous evidential profile. To address this challenge, I evaluate three frameworks: EBM, EBM+, and Bayesianism. EBM is limited by its evidential monism and undervaluation of mechanistic reasoning and, therefore, is unable to effectively evaluate a treatment like PBT. EBM+, inspired by the Russo-Williamson Thesis, recognises the need for both mechan-

istic and statistical evidence but lacks procedural clarity for integrating them. I argue that a Bayesian framework offers a superior solution. By allowing mechanistic evidence to inform priors and statistical evidence to update beliefs via likelihoods, Bayesianism provides a principled and transparent method for integrating diverse evidence under uncertainty. It thereby offers a more coherent foundation for causal inference in MBM and a more appropriate evidential standard for regulators and insurers evaluating technologies such as PBT.

A Significance Test for Causal Inference

Zhitao Zhang
Marche Polytechnic University
zhangzhitao743@gmail.com

Alexander Gebharter Marche Polytechnic University
alexander.gebharter@gmail.com

Graphical causal models have become very powerful tools for causation-related debates. It is usually assumed that causes are distinct from their effects (Lewis 2000, p. 78). In order to avoid possible conceptual problems and unwanted consequences, typical causal models do not feature non-causal relations such as constitution or supervenience. Kim (2000, 2005), for example, argued for the controversial view that mental causation is excluded in the presence of supervenience and physical causation. Mixing variables at different levels in one model also triggers confusion about what counts as an intervention or manipulation (e.g., Baumgartner and Gebharter 2016, Woodward 2015, Baumgartner 2018).

Despite these problems arising from allowing non-causal dependencies in causal models, both the scientific and the philosophical communities see a need for representing systems featuring not only causal, but also other relationships. For example, Andersen (2024) pointed out that in fields like biomedical physiology, it is disproportionately likely that different groups will simultaneously study overlapping but non-identical parts of the same phenomenon (i.e., the kidney both at fine and coarse scales). In philosophy of science, both the debate on mental causation (e.g., Kroedel 2024) and that on inter-level causation in mechanistic explanation (e.g., Kistler 2009, Craver et.al. 2021) are committed to accounting for causation at different levels in one way or another.

In this paper, we do not take sides in ontological debates about whether higher- or inter-level causation exists or how one should account for it. Given the need for a model that incorporates causality at different levels, we develop an account aimed at aiding scientific practice rather than deciding a metaphysical debate.

Our account takes inspiration from Craver et al.'s (2021) comment that constitution relationships are best understood as a constraint. We build models that represent constitutive relevance relations as synchronization conditions between different levels, and treat causal connections as relations among variables within particular levels that are constrained by synchronization relations. We argue that this constraint fits scientific practice by only restricting the range of certain variables given information about the variables at other levels.

We demonstrate several merits of the account. Firstly, inter-level inference under observation works smoothly. Secondly, the account allows for computing the effects of interventions not only within, but also across levels of organization. In our models, higher-level interventions largely preserve the underlying lower-level causal structure, which closely reflects scientific practice in biology and medicine. Thirdly, the account faithfully represents and explains the process (i.e., experiments) by which scientists infer

constitutive relevance. It also provides a formal representation of Craver's improved version of the Mutual Manipulability criterion. Finally, the account can in principle be generalized for any type of synchronization among models, making it flexible enough to formally engage other non-causal dependencies such as statistical correlations, physical laws, etc. into causal models. This provides a way to study causation more realistically.

References

Andersen, H. (2024). Why adoption of causal modelling methods requires some metaphysics. In F. Russo & P. Illari (Eds.), The Routledge handbook of causality and causal methods. New York: Routledge.

Baumgartner, M. & Gebharter, A. (2016). Constitutive relevance, mutual manipulability, and fat- handedness. British Journal for the Philosophy of Science, 67(3), 731-756.

Baumgartner, M. (2018). The inherent empirical underdetermination of mental causation. , 96 (2), 335–350.

Craver, C. F., Glennan, S. & Povich, M. (2021). Constitutive Relevance & Mutual Manipulability Revisited. Synthese, 199:8807–8828.

Gebharter, A. (2017). Uncovering constitutive relevance relations in mechanisms. Philosophical studies. 174(11), 2645-2666.

Kim, J. (2000). Mind in a physical world. MIT Press.

Kim, J. (2005). Physicalism, or something near enough. Princeton University Press.

Kistler, M. (2009). Mechanisms and downward causation. Philosophical Phycology, 22(5), 595-609.

Kroedel, T. (2024). How not to intervene on mental causes. Philosophical Studies, 181:2479–2497.

Lewis, D. (2000). Causation as influence. The Journal of Philosophy, 97(4), 182-197.

Woodward, J. F. (2015). Interventionism and causal exclusion. Philosophy and Phenomenological Research, 91 (2), 303–347.