




FROM THE INSIDE



Falsifiability in medicine: what clinicians can learn from Karl Popper

Shaurya Taran^{1*} , Neill K. J. Adhikari^{2,3}  and Eddy Fan^{3,4} 

© 2021 Springer-Verlag GmbH Germany, part of Springer Nature, corrected publication 2021

This isn't right. It's not even wrong!
-Wolfgang Pauli.

In the early twentieth century, the philosopher Karl Popper became intrigued by a basic question in the philosophy of science: how does one distinguish true science from non-science? For Popper, the distinction hinged on the essential ingredient of falsifiability [1]. True science was falsifiable: it could be proven incorrect by an experiment that contradicted its predictions. Non-science, on the other hand, was unfalsifiable: it made no predictions that could be disproven by experimental methods. Popper highlighted the difference using Einstein's theory of general relativity and Freud's theory of psychoanalysis as examples. Einstein's theory inferred specific claims about the natural world. It invited experimentation and set itself up to be either *corroborated* or *falsified* by experiments. By contrast, Freud's theory used observations to posit a general theory about human nature, but for a given patient, it made no specific predictions. Since no experiment could be put forth to contradict it, Popper regarded it as unfalsifiable. This concept was also famously highlighted by the physicist Wolfgang Pauli, who, when asked to review a paper that he deemed to be unfalsifiable, lamented, "This isn't right. It's not even wrong!"

Although Popper's theories about the essence of science have been challenged, the core concept of falsifiability to assess whether a claim is scientific has endured. For clinicians, Popper's notion might be used to evaluate new ideas and decide how much weight to give them. This

framework is important because clinicians today, more than ever before, face an explosion of ideas of variable quality, making it difficult to know where to place one's trust. There are so many provocative conjectures and expert opinions being circulated that the task of carefully assessing their credibility has never been more necessary. The current coronavirus disease 2019 (COVID-19) pandemic has offered several examples of conjectures widely adopted without rigorous evaluation, sometimes leading to patient harm. In light of this chaos, what lessons might Popper's notion of falsifiability hold for clinicians, and how can these lessons help clinicians become better judges of science?

First, there is a useful difference between conjectures and theories. Conjectures about medicine may stem from uncontrolled clinical observations in patients, physiological experiments, or animal models—sources of evidence whose trustworthiness is often downgraded because of high risk of bias, indirectness to actual clinical problems in patients, or imprecision of the estimated treatment effect [2]. On their own, conjectures may be difficult to falsify, because they generate no new predictions, or because they rely on a series of linked conjectures to generate testable hypotheses. By contrast, tentative theories emerge from a preponderance of data from internally valid studies whose results point in a consistent direction. Theories can be corroborated or falsified by high-quality tests, such as randomized clinical trials (see Fig. 1 for additional differences between conjectures and theories). This distinction bears reminding in the context of the ongoing pandemic, where conjectures have often shifted clinical practice in a manner out of proportion to the certainty of evidence. Consider the example of hydroxychloroquine, which was touted as a breakthrough treatment for COVID-19 on the basis of *in vitro* studies demonstrating anti-viral activity [3] and studies of fewer than 30 patients that showed reduced viral nasopharyngeal

*Correspondence: Shaurya.taran@mail.utoronto.ca

¹ Interdepartmental Division of Critical Care Medicine, Li Ka Shing Knowledge Institute, University of Toronto, 204 Victoria Street, 4th floor room 411, Toronto, ON M5B 1T8, Canada

Full author information is available at the end of the article



Conjectures	Theories
Arise from uncontrolled observations and preliminary supporting evidence which is often incomplete	Arise from a series of internally valid studies whose conclusions are concordant
Predictions are overly general or lack a defined context, making them difficult to falsify by experimental methods	Predictions are specific and applicable in a well-defined <i>a priori</i> context, setting up the design of a fair test or study
High risk of bias and other problems raise reasonable doubts about the strength of evidence	Low risk of bias and other strengths provide reassurance that the evidence is high-certainty
	
Findings should be regarded as invitations to further study; may be appropriate for clinical application, but usually in limited circumstances (e.g. a ‘parachute effect’)	Findings can be used to guide clinical management while being cautious about residual uncertainty

Fig. 1 Differences between conjectures and theories

carriage [4]. In France, where one of these studies was performed, prescriptions of hydroxychloroquine surged. Notwithstanding the fact that multiple subsequent clinical trials showed no benefit, and possibly increased harm, associated with hydroxychloroquine, it is unsettling that early conjectures were rapidly adopted across the medical community before a theory could emerge and a fuller understanding of its risks and benefits be appreciated. When evaluating a new idea, Popper thus encourages clinicians to ask the following questions: does the available body of knowledge describe a conjecture or a theory? Have the data been evaluated by a high-quality test? And if not, could it theoretically be corroborated or falsified by such a test in the future, provided that there remains equipoise about the intervention [5]? These questions can help clinicians decide if an idea is in a rudimentary or more advanced phase of development—and accordingly, whether it deserves further testing or is ready for application.

Second, although falsifiability is a binary concept (an idea is either falsifiable or it isn't), theories are more complex: they might be completely true under some conditions, completely untrue under others, or partially true depending on which aspects are considered.

When interpreting new studies, it is therefore important to appreciate these nuances and resist the tendency to oversimplify. Consider the example of dexamethasone for the treatment of patients with COVID-19. In June 2020, preliminary results from the RECOVERY trial were released, which showed that in mechanically ventilated patients with COVID-19, treatment with dexamethasone resulted in a 11.7% absolute risk reduction in 28-day mortality compared to usual care [6]. Among non-ventilated patients receiving oxygen, dexamethasone resulted in a less pronounced, but significant, absolute risk reduction in 28-day mortality of 3.5%. However, in the group of patients not receiving oxygen or mechanical ventilation, no mortality benefit with dexamethasone was observed. These landmark findings were rapidly communicated throughout the lay press, often with the simple bottom-line message that dexamethasone saved lives, without further elaboration into the groups most likely to benefit or not benefit at all. In reality, the role of corticosteroids in COVID-19 is far more nuanced, with a differential response depending on disease severity. To ignore such complexities—as has been done by politicians and policymakers on various matters throughout this pandemic—misrepresents the truth and propagates

misunderstandings. It is therefore always worth asking what the study corroborated or falsified before making a judgment about the theory on the whole.

Third, theories which have accumulated a wealth of evidence, generally over a long period of time and by many examiners, have withstood Popper's falsifiability test, perhaps many times over: they are the best approximators of real truth. In Bayesian terms, ideas with a long history of consistent messaging offer a reliable set of priors to understand and evaluate new evidence. This concept could be remembered, for example, when one is confronted by ideas that do not fit with established priors. Consider the example regarding the existence of "H" and "L" phenotypes of COVID-19-related acute respiratory distress syndrome (ARDS) [7]. Investigators hypothesized that there are two distinct COVID-19 ARDS phenotypes (with a spectrum in between) that mandate different approaches to mechanical ventilation. They suggested that not identifying the correct phenotype might lead to selection of the wrong ventilation approach and patient harm. These hypothetical phenotypes were not previously demonstrated to exist, but more importantly, the ventilation strategy proposed for "L"-type patients (i.e., use of high tidal volumes) contradicts decades of ARDS research, which has shown multiple benefits in favor of lower tidal volume ventilation [8]. This is not to say that the "H" vs "L" phenotype conjecture should not be subject to further testing—but rather, it is difficult to justify abandoning a robust set of priors when a single new concept challenges them. As Popper might argue, the preponderance of existing evidence on an idea should guide clinicians in deciding where to place their trust while awaiting the results of additional investigations.

Popper applied the notion of falsifiability to distinguish between non-science and science. Clinicians might apply the same notion to understand and evaluate new ideas. This process entails three key considerations. First, conjectures should be seen as invitations to design further studies to evaluate them. While many conceptually interesting new conjectures move their fields in a novel direction, they often still require confirmation by high-quality studies, and their application to patient care might be premature (or even harmful) before such validation occurs. Second, for theories that have been apparently corroborated or falsified by high-quality tests, it is still worth asking: what did the tests actually show? Did they prove or disprove the entire theory, or only some aspect of it? This interpretation of data, formalized in the GRADE system of assessing certainty of evidence [2], should guide the application of research findings into the real world. Finally, when evaluating an idea for which

there is existing knowledge, it is worth placing the idea in its available context. Ample time also provides ample opportunities for falsification. Ideas that have withstood Popper's test are probably robust—and more likely to be "true" as compared to the "new truths" with which we are all nowadays regularly confronted.

Author details

¹ Interdepartmental Division of Critical Care Medicine, Li Ka Shing Knowledge Institute, University of Toronto, 204 Victoria Street, 4th floor room 411, Toronto, ON M5B 1T8, Canada. ² Sunnybrook Health Sciences Centre, Toronto, ON, Canada. ³ Institute for Health Policy, Management, and Evaluation, University of Toronto, Toronto, ON, Canada. ⁴ Toronto General Hospital, University Health Network, Toronto, ON, Canada.

Author contributions

All authors were involved in article conception, manuscript preparation, and critical revisions.

Declarations

Conflicts of interest

All authors declare that they have no competing conflicts of interest.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 3 April 2021 Accepted: 7 May 2021

Published online: 22 May 2021

References

1. Popper K (1968) *The Logic of Scientific Discovery*, 7th edn. Harper & Row, New York
2. Guyatt GH, Oxman AD, Vist GE et al (2008) GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 336:924–926
3. Yao X, Ye F, Zhang M et al (2020) In vitro antiviral activity and projection of optimized dosing design of hydroxychloroquine for the treatment of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). *Clin Infect Dis* 71:732–739
4. Gautret P, Lagier JC, Parola P et al (2020) Hydroxychloroquine and azithromycin as a treatment of COVID-19: results of an open-label non-randomized clinical trial. *Int J Antimicrob Agents* 56:105949
5. Yeh RW, Valsdottir LR, Yeh MW et al (2018) Parachute use to prevent death and major trauma when jumping from aircraft: randomized controlled trial. *BMJ* 363:5094
6. Horby P, Lim WS, Emberson J, et al. Effect of Dexamethasone in Hospitalized Patients with COVID-19 – Preliminary Report. *medRxiv* 2020:2020.06.22.20137273
7. Gattinoni L, Chiumello D, Caironi P et al (2020) COVID-19 pneumonia: different respiratory treatments for different phenotypes? *Intensive Care Med* 46:1099–1102
8. Fan E, Del Sorbo L, Goligher EC et al (2017) An official american thoracic society/european society of intensive care medicine/society of critical care medicine clinical practice guideline: mechanical ventilation in adult patients with acute respiratory distress syndrome. *Am J Respir Crit Care Med* 195:1253–1263