Commentary

The Call for Progress in Evidence-Based Medicine

Franz Porzsolt[1]

1. Private Research Institute of Clinical Economics, Germany

Randomized controlled trials (RCTs) are foundational in evidence-based medicine but are often limited by their reliance on controlled conditions and highly selective patient populations, which may not reflect real-world healthcare outcomes. This commentary highlights the value of pragmatic controlled trials (PCTs) as a complementary method for evaluating real-world effectiveness (RWE). By employing Bayesian statistical approaches, PCTs enable comprehensive assessment of healthcare interventions under everyday conditions, bridging the gap between experimental efficacy and practical applicability. A three-dimensional strategy is proposed, addressing efficacy, effectiveness, and value to improve the relevance and generalizability of healthcare research. The commentary also emphasizes the need for endpoint-specific risk profiling to optimize patient outcomes and decision-making. Integrating PCTs into evidence-based medicine provides a robust framework for advancing healthcare efficiency and delivering patient-centered care in diverse populations.

Introduction

Four international cardiology societies (European Society of Cardiology, American Heart Association, American College of Cardiology, World Heart Federation) issued a joint statement proposing a modification of randomized controlled trials (RCT)[1]. This modification is justified due to increased administrative requirements and financial burdens, as well as a disproportionately low information gain from conventional RCTs. In a 'joint opinion', the design of an adaptive platform study is proposed instead of traditional RCTs[2] because promising results could be achieved by this study design in different studies[3][4][5].

We agree with the Joint Opinion Group's call for a necessary optimization of the standards for gaining knowledge in the healthcare system and contribute our experience we gained while developing the

pragmatic controlled trial (PCT).

Background

The need to develop a specific method for the detection of non-experimental care effects arose in the late 1980s. One of us, a young oncologist, noticed that treatment successes in patients at our university hospital were inferior compared to published oncology reports. Nearly a decade passed before a simple idea could plausibly explain the difference. We observed effects in our hospital that occur in everyday care (real-world effectiveness), whereas journals reported data were almost exclusively generated in experimental studies under strictly controlled conditions. Although the scientific literature differentiated between "efficacy" and "effectiveness" very early[6][7][8], convincing methods for the undistorted detection of results under non-experimental conditions were not yet available. We did not succeed in formally describing the difference between expected and observed results until much later[9].

Our research in evidence-based medicine taught us Sir Archibald Cochrane's and Sir Austin Bradford Hill's three essential questions to ask before implementing an innovation in the healthcare system: "Can it work? Does it work? Is it worth it?"[10]. Cooperation with teachers and students in the "hochschule für gestaltung (hfg)" (Ulm school of design) taught us the rule "Form Follows Function (FFF)"generated by American designers and architects[11]. As citizens of Ulm, we are familiar with many of Albert Einstein's (born in 1879 in Ulm) statements, which pointed out that problems cannot be solved by the mindset that caused them. The recommendations of the British epidemiologists and input by the American designers and the German physicist facilitated the development of a three-dimensional strategy for the evaluation of healthcare performance.

The three-dimensional strategy

The concept of the three-dimensional strategy is based on the three Cochrane-Hill questions. The answer to the first question, "Can it work?", requires proof of the effective principle (proof of principle, PoP). The second question, "Does it work?" can be answered by demonstrating real-world effectiveness (RWE). The third answer describes the perceived value (Val) of healthcare services from an individual and a societal perspective. Efficacy and effectiveness depend on objective judgments, whereas the description of value is a subjective but essential judgment. If the proof of efficacy is

supplemented by the proof of real-world effectiveness, not only effective interventions can be identified. In addition, it will also be possible to describe the endpoint-specific risk profiles of patients who can be successfully treated with effective intervention. This classification of the successfully treatable subpopulation will help to significantly increase care efficiency. If different therapies achieve identical results in patients with identical risk profiles, this knowledge will also contribute to the development of new care strategies.

The criteria for distinguishing efficacy and effectiveness

Functions and forms of the three healthcare conditions

Patient care can be performed under three different conditions, under the non-structured, natural conditions of everyday care or under two different types of a structured study. The two structured study conditions are the strictly controlled experimental or interventional study, the randomized controlled trial (RCT), or the non-experimental or observational study, the pragmatic controlled trial (PCT), which is based on the principle of Bayes' statistics. In a PCT, each patient is cared for under non-structured, natural everyday care conditions, but evaluated under structured conditions by applying Bayesian statistics. The advantage of Bayesian statistics over randomization is the ability to apply statistical methods under the non-structured conditions of everyday care without altering these natural conditions.

The Bayesian method requires documentation of the intervention (therapy) and all individual risk factors that could affect any of the measured PCT endpoints. Based on the categorization of therapies and individual risk profiles, each patient can be assigned to the appropriate risk class with respect to each of the measured PCT endpoints[12][13][14][15]. Although this accurate risk classification requires large numbers of cases, it allows comparison of patients assigned to an individual risk class with respect to each measured endpoint and the intervention applied. An RCT only ensures that the risk profiles are equally distributed within the study arms. Therefore, all patient risk profiles that were not excluded in the RCT are represented in each study arm. Only a limited number of different therapies can be studied in an RCT, which significantly limits the applicability of results in everyday care. We compiled the detailed differentiation of the three healthcare conditions based on two functional and twelve formal (structural) criteria[16].

Criteria	Efficacy	Effectiveness
Functions	Care of subjects under **structured** conditions of an experimental study. **Proof of principle (PoP)** under structured conditions of an experimental study (e.g. RCT)	Care of subjects under **not structured** conditions of everyday care Demonstration of **real-world effectiveness (RWE)** under structured conditions of an observational study (e.g. PCT).
Forms / structures	No agreement between the 12 criteria of an experimental study (RCT) with the criteria under the conditions of everyday care.	Agreement of six of the 12 criteria with the criteria of experimental studies and of four criteria with the criteria of everyday care

Table 1. Functions and forms / structures of efficacy and effectiveness[15]

Causes, consequences, and a possible solution to the terminology conflict

The challenge to assessing healthcare outcomes in three dimensions[10], which has been unresolved for 80 years, is based on a terminology conflict involving differentiating efficacy from effectiveness[14]. The most likely cause of this terminology conflict is the lack of proposed alternative solutions. Therefore, the RCT has hitherto been considered the only valid method to measure the effects of healthcare.

The consequence is that most clinical decisions concerning guidelines, patient treatment, and court verdicts are decided by the experimentally driven proof of principle and not by its suitability for everyday use. These decisions are based on highly selected patient populations and often on surrogates rather than real endpoints. As a result of these two "idealized measurement conditions", the successes achievable under everyday conditions are significantly overestimated. This overestimation can be avoided by using PCTs.

Differentiation among the three outcome dimensions and the three healthcare conditions

Table 2 describes the three outcome dimensions of proof of principle (PoP), real-world effectiveness (RWE), and value (Val) from the perspectives of clinical research, health-services research, and

economic research. Each addresses different questions, different healthcare conditions, different study types, and different methods of achieving evidence[13].

Perspective	Question	Answer	Health care condition	Type of study	Method
Clinical research	Can it work?	Objective confirmation of proof of principle (PoP) or efficacy	Experimental study condition (ESC)	Interventional study	Randomized controlled trial (RCT)
Health services research	Does it work?	Objective confirmation of real-world effectiveness (RWE)	Real-world condition including with systematic evaluation of outcomes	Pragmatic / observational study	Pragmatic controlled trial (PCT)
Economic research	Is it worth it?	Subjective confirmation of value (Val)	Real-world condition without systematic evaluation of outcomes	Complete economic analysis	Cost-effectiveness analysis (CEA)

Table 2. Answering the three Cochrane–Hill questions from the perspectives of clinical research, health services research, and economic research (modified from[12]).

Healthcare services have hitherto been evaluated by objective evidence of PoP, where final policy decisions are almost always made on the basis of the subjective estimation of value[17][18]. Due to the measurability of RWE, the final subjective decision can be supported by data that are closer to the objectifiable value of a healthcare service than the proof of the PoP.

Our assumption that we doctors make arbitrary decisions in everyday care is probably incorrect. Every doctor makes an implicit effort to adapt his strategy to the individual risk profile of his patient. However, this strategy has not yet been standardized[16]. In everyday health-care, we can distinguish three conditions under which healthcare services are offered. The experimental conditions require to conduct a structured RCT. These studies are well known. We recently described two functional and twelve formal criteria to distinguish these conditions from healthcare conditions under the non-structured conditions of everyday care[17][18]. In PCTs, which detect effects induced under everyday conditions, six of the twelve formal criteria are consistent with the criteria used in experimental studies, and four criteria are consistent with the criteria used in the non-structured conditions of everyday care. Two criteria of the PCT differ from all criteria of an experimental study and the criteria of everyday care[18].

Importance of the study question, the study conditions (including selection criteria), and the interpretations

Under the title "Front-end-processor", we present data suggesting that every scientific question may be developed in four steps. If these four steps do not correspond exactly in terms of content, the risk of answering the scientific question incorrectly will increase[18]. We also address the necessary congruence of the forms (structures) and functions of research methods. Experimental methods cannot be used for the analysis of outcomes of (non-experimental) care as usual[18]. The reproduction of reported outcomes may be impossible unless the risk profiles of he investigated patients were known[19]. Scientists and policy makers use the same (experimental) data to "make" and to "take" different types of decisions[20]. Using the example of breast cancer screening we show that the same data – analyzed in different ways – can quantify both the objective risks and the subjective perception of objective risks[18][21]. More attention should be paid to the study conditions and the selection criteria because, without their descriptions, it is difficult to draw conclusions about the scope of the results collected[20]. Both the chosen study conditions and the chosen selection criteria influence the study results via direct and indirect effects. Direct effects concern the exact formulation of the study objective, the results obtained, and their interpretability. Indirect effects of a healthcare study define its scope. Since defined selection criteria in studies that are included in meta-analyses and/or confirm similar effects are often only vaguely defined, a sharp delineation of the scope of clinical trial results cannot always be deduced[17][18].

The outcome dimension (PoP, RWE, or Val) also needs to be defined because the demonstration of each of these dimensions requires different methods and strategies. The correspondence between the precise question to be answered and the choice of the most appropriate endpoints to achieve this answer is crucial for the quality of the responses obtained. The more differentiated the inclusion criteria of a study, the more uniform (but also smaller) the population studied and the more likely it will be possible to obtain consistent results.

Understanding the significant differences between inclusion and exclusion criteria is important. Inclusion criteria are required for any form of health-related study as opposed to exclusion criteria. Exclusion criteria exist only in experimental studies, but not in studies describing everyday healthcare, like the PCT or studies describing the subjectively perceived added value of a healthcare service or health-related quality of life[9][19]. Both exclusion and inclusion criteria, depend on the

study question. The function of exclusion criteria, however, is to describe all subjects who exhibit any of the confounding factors that may bias the measurement of the primary endpoint of an experimental study. It should be noted that exclusion criteria only address treatment with the study medication and selection of the study population, not treatment of the excluded subjects with all other therapies. In other words, exclusion criteria protect the evidence of PoP from bias, but compromise the evidence of RWE because the risk profiles of patients investigated in experimental studies will barely meet the conditions of care as usual.

Deriving the consequences

The results of all decision-relevant studies, whether experimental or pragmatic, have often been applied without exact consideration of the selection (inclusion and exclusion) criteria[19]. When evaluating services in the health care system, we scientists should pay close attention to the patient population being examined. When highly selected (experimental) populations have been investigated, recommendations for everyday care can hardly be derived. Knowledge of our patients' endpoint-specific risk profiles can improve the quality and efficiency of healthcare. The benefits of analyzing endpoint-specific risk profiles should be particularly evident in very large study groups, as the expected variance of these profiles will be rather high. For a systematic analysis, comparable risk profiles are to be stratified into similar (high, intermediate, low) endpoint-specific risk classes. This requirement can only be met if this classification is carried out according to jointly defined criteria. These considerations presuppose the willingness to cooperate in very large projects.

Discussion and summary

Evidence of fitness for daily use should be demonstrated for all interventions applied in healthcare. The shift of focus from PoP to RWE can be justified by the following:

- RCT studies only involve a highly selected patient population in which the major risk factors affecting the measured primary endpoint have been eliminated by exclusion criteria. Exclusion criteria are not applied in a PCT because they would not accurately describe the population receiving healthcare under everyday conditions.
- An RCT limits the choice of healthcare options to the few interventions that can be compared and interpreted in an RCT. The PCT does not limit the choice of healthcare options. Each participant

chooses the intervention expected to produce the optimal outcomes for the individual patient. This produces the healthcare conditions applied under everyday medical practice.

- An RCT is expected to ensure the equal distribution of all risk factors not excluded in the study populations. This, however, can hardly be confirmed because the size of the studied population depends on a large number of variables, like the number of risk factors, their effect sizes, and their interrelationships. The smaller the population studied in an RCT, the greater the danger that not all risks will be equally distributed in the randomized groups.

Progress in health care can be achieved step by step. The supplied patents will only notice that considerably more data is collected than before, but that the supply will remain unchanged for the time being. The advanced data collection will require several basis steps.

1. Selection of the clinical health problem to be analyzed.

2. Definition of the targeted endpoints of care in advance.

3. Definition of the potential risk factors of the patients that may impair the achievement of these endpoints, i.e. the "endpoint-specific risk lists (ESRLs)".

4. Based on these ESRLs, clinical expert teams can form different endpoint-specific risk classes (ESRCs; high, intermediate, low).

5. To evaluate the care outcomes, each patient treated is assigned to a defined ESRC (high, intermediate, low) for each measured endpoint. The methods of AI enable this complex data assessment and collection, which includes not only the risk profile of the patient but also a classification of the therapeutic measures. Usually, multiple health problems require multiple therapies in parallel in the majority of patients[12][13][14][15][16][17][18].

The necessary increase in data collection may be perceived by doctors as a similar burden as the demand for randomization 30 years ago. Nevertheless, there will be a significant difference because the assessment of the risk profile will seem plausible for patients and physicians and, unlike randomization, will not affect the relationship between physician and patient.

A change in our traditional way of thinking is necessary to accept that the proof of everyday suitability of healthcare services, i.e., the new field of healthcare-services research, requires two different healthcare conditions (twin method)[15]: Care must be provided under the non-structured everyday conditions of 'natural chaos' prevailing in patient care, while the evaluation of healthcare outcomes requires precisely structured tools, like Bayesian statistics, with no reciprocal influence between these

two methods, the care as usual and the method used for the analysis of the data. This comment should appeal to colleagues who share our concern that the uncritical interpretation of the results of experimental RCTs could affect the financial viability of our health systems.

Several scientists doubted the results of the RCTs[22][23]. However, it is possible that the method of the RCT itself is not the problem. The randomization of patients (in contrast to well defined objects) requires compliance with specific framework conditions, e.g. the exclusion of certain risks. Consequently, the interpretation of the results of an RCT will only be valid if the limitations defined by the framework conditions are actually taken into account. Otherwise, the effects that can actually be achieved will be overestimated. RCTs cannot provide detailed data to derive new care concepts because the risk profiles of the patients treated will be too different. This assumption suggests that PCTs should not be carried out regionally, but at the national level. Without taking into account the national perspective, there is a risk of overlooking risks that are specific to certain regions. In other words, the orientation of care towards ESRPs is probably more complex than originally expected.

This commentary does not claim to discuss all the details of the detailed assessment of supply effects. However, it pursues the goal of increasing interest in proving RWE. Without this proof, it will hardly be possible to prove the efficiency of health services.

Acknowledgements

References

1. ^Bowman L, Weidinger F, Albert MA, Fry ETA, Pinto FJ. "Randomized Trials Fit for the 21st Century. A Joint Opinion from the European Society of Cardiology, American Heart Association, American College of

Cardiology, and the World Heart Federation." Global Heart. 17 (1): 85. doi:10.5334/gh.1178.

2. ^*Park JJH, Detry MA, Murthy S, Guyatt G, Mills EJ. "How to use and interpret the results of a platform tri al: users' guide to the medical literature." JAMA. 327 (1): 67-74. doi:10.1001/jama.2021.22507.*

3. ^*Gordon AC, Angus DC, Derde LPG. "Interleukin-6 Receptor Antagonists in Critically Ill Patients with Co vid-19. Reply." N Engl J Med. 2021 Sep 16;385(12):1147-1149. doi: 10.1056/NEJMc2108482. PMID: 3440 7335.*

4. ^*Writing Committee for the REMAP-CAP Investigators; Higgins AM, Berry LR, Lorenzi E, Murthy S, McQ uilten Z, Mouncey PR, Al-Beidh F, Annane D, Arabi YM, Beane A, van Bentum-Puijk W, Bhimani Z, Bont en MJM, Bradbury CA, Brunkhorst FM, Burrell A, Buzgau A, Buxton M, Charles WN, Cove M, Detry MA, E stcourt LJ, Fagbodun EO, Fitzgerald M, Girard TD, Goligher EC, Goossens H, Haniffa R, Hills T, Horvat C M, Huang DT, Ichihara N, Lamontagne F, Marshall JC, McAuley DF, McGlothlin A, McGuinness SP, McVe rry BJ, Neal MD, Nichol AD, Parke RL, Parker JC, Parry-Billings K, Peters SEC, Reyes LF, Rowan KM, Sait o H, Santos MS, Saunders CT, Serpa-Neto A, Seymour CW, Shankar-Hari M, Stronach LM, Turgeon AF, Turner AM, van de Veerdonk FL, Zarychanski R, Green C, Lewis RJ, Angus DC, McArthur CJ, Berry S, Derd e LPG, Gordon AC, Webb SA, Lawler PR. "Long-term (180-Day) Outcomes in Critically Ill Patients With COVID-19 in the REMAP-CAP Randomized Clinical Trial." JAMA. 2023 Jan 3;329(1):39-51. doi: 10.1001/ jama.2022.23257. PMID: 36525245; PMCID: PMC9857594.*

5. ^*Barnett ML, Sax PE. "Long-term Follow-up After Critical COVID-19: REMAP-CAP Revisited." JAMA. 2 023 Jan 3;329(1):25-27. doi: 10.1001/jama.2022.23700. PMID: 36525259.*

6. ^*Schwartz D, Lellouch J. "Explanatory and pragmatic attitudes in therapeutic trials." J. Chron. Dis 1967; 20: 637-648 (1967).*

7. ^*Grimes DA, Schulz KF. "An overview of clinical research: the lay of the land." Lancet 2002:359; 57-61.*

8. ^*Thiese MS. "Observational and interventional study design types; an overview." Biochem Med 2014: 2 4(2); 199-210 (2014). http://dx.doi.org/10.11613/BM.2014.022.*

9. a, b*Wiedemann F, Porzsolt F. "Measuring Health-Related Quality of Life in Randomised Controlled Tria ls: Expected and Reported Results Do Not Match." Pragmat. Obs. Res. 2022 Apr. 11;13:9-16. doi: 10.2147/ POR.S350165. PMID: 35431592; PMCID9012498.*

10. a, b*Haynes B. "Can it work? Does it work? Is it worth it? The testing of healthcare interventions is evolvi ng." BMJ. 1999 Sep 11;319(7211):652-3. doi: 10.1136/bmj.319.7211.652. PMID: 10480802; PMCID: PMC111 6525.*

11. <u>^</u>*Sullivan, L.H. "The tall office building artistically considered." Lippincott's Magazine 1896;57: 403-40 9. Reprinted in Inland Architect and News Record 27 (May 1896), pp. 32-34; Western Architect 31 (Janua ry 1922), pp. 3-11; published as "Form and Function Artistically Considered" The Craftsman 8 (July 190 5), pp. 453-58. https://www.giarts.org/article/tall-office-building-artistically-considered-excerpt. Las t access oct 22, 2019*

12. <u>a</u>, <u>b</u>, <u>c</u>*Porzsolt F, Eisemann M, Habs M, Wyer P. "Form Follows Function: Pragmatic Controlled Trials (PC Ts) have to answer different questions and require different designs than Randomized Controlled Trials (RCTs)." J Publ Health 2013;21:307-313. DOI 10.1007/s10389. https://www.ncbi.nlm.nih.gov/pmc/article s/PMC3655212/*

13. <u>a</u>, <u>b</u>, <u>c</u>*Porzsolt F, Rocha NG, Toledo-Arruda AC, Thomaz TG, Moraes C, Bessa-Guerra TR, Leão M, Migow ski A, Araujo de Silva AR, Weiss C. "Efficacy and Effectiveness Trials Have Different Goals, Use Different Tools, and Generate Different Messages." Pragmatic and Observational Research 2015;6:47-54. DOI htt p://dx.doi.org/10.2147/POR.S89946*

14. <u>a</u>, <u>b</u>, <u>c</u>*Porzsolt F, Wiedemann F, Phlippen M, Weiss C, Weiss M, Schmaling K, Kaplan RM. "The terminolo gy conflict on efficacy and effectiveness in healthcare." J Comp Eff Res. 2020;9:1171-1178. doi: 10.2217/ce r-2020-0149.*

15. <u>a</u>, <u>b</u>, <u>c</u>, <u>d</u>*Porzsolt F, Weiss Ch, Weiss M. "Covid-19: Hybridmethode zum Nachweis der Real-World Effecti veness unter Alltagsbedingungen [Covid-19: Twin method for demonstration of real-world effectivenes s (RWE) under the conditions of day-to-day care]." Das Gesundheitswesen 2022;84:1-4. DOI: 10.1055/a -1819-6237. ISSN 0941-3790.*

16. <u>a</u>, <u>b</u>, <u>c</u>*Porzsolt F, Weiss Ch, Weiss M. "An Evidence-Based Hypothesis: Doctors Do Not Make Decisions Ra ndomly but Based on Individual Patient's Risk Profiles." Preprint from Qeios, 21 May 2024 https://doi.o rg/10.32388/no2184 PPR: PPR856245*

17. <u>a</u>, <u>b</u>, <u>c</u>, <u>d</u>*Porzsolt F, Weiss M, Weiss Ch. "Applying the Rule of Designers and Architects "Form Follows Fun ction (FFF)" Can Reduce Misinterpretations and Methodical Shortcomings in Healthcare." Trends Gen Med. 2024; 2(1): 1-7. ISSN: 2996-3893*

18. <u>a</u>, <u>b</u>, <u>c</u>, <u>d</u>, <u>e</u>, <u>f</u>, <u>g</u>, <u>h</u>*Porzsolt F, Phlippen MS, Legrum P, Weiss M. "The Front-End Processor Developed by En gineers — A Useful Tool for Describing the Quality and Quantity of Progress in Healthcare." Qeios ID: 8P WWZD · https://doi.org/10.32388/8PWWZD*

19. <u>a</u>, <u>b</u>, <u>c</u>*Porzsolt F, Wiedemann F, Becker SI, Rhoads CJ. "Inclusion and exclusion criteria and the problem o f describing homogeneity of study populations in clinical trials." BMJ Evidence-Based Medicine 2019; d*

oi:10.1136/bmjebm-2018-111115

20. a, b Gray, JA Muir. "Evidence based policy making." BMJ 2004; 329: 988-989.

21. ^ Porzsolt F, Pfuhl G, Kaplan RM, Eisemann M. "Covid-19 pandemic lessons: Uncritical communication of test results can induce more harm than benefit and raises questions on standardized quality criteria f or communication and liability." Health Psychology and Behavioral Medicine 2021;9:818-829. DOI 10.1 080/21642850.2021.1979407

22. ^ Krauss A. "Why all randomised controlled trials produce biased results." Annals of Medicine 2018;50:31 2-322. https://doi.org/10.1080/07853890.2018.1453233

23. ^ Jureidini J, McHenry LB. "The illusion of evidence based medicine." BMJ. 2022 Mar 16;376:o702. doi: 1 0.1136/bmj.o702. PMID: 35296456.

Declarations