



Contents lists available at ScienceDirect

European Journal of Internal Medicine

journal homepage: www.elsevier.com/locate/ejim

Learning in times of stress: Lessons from COVID-19 that will last throughout this century

Paula Jiménez-Fonseca^{a,*}, Ramón Salazar^b, Vicent Valentí^c, Alberto Carmona-Bayonas^d, Giancarlo Agnelli^e

^a Medical Oncology Department, Hospital Universitario Central de Asturias, Oviedo, Spain

^b Medical Oncology Department, Oncobell Program, IDIBELL Institut Català d'Oncologia, Hospital Duran i Reynals, CIBERONC, Barcelona, Spain

^c Medical Oncology Department, Hospital Del Vendrell, El Vendrell, Tarragona, Spain

^d Hematology and Medical Oncology Department, Hospital Universitario Morales Meseguer, UMU, IMIB, Murcia, Spain

^e Internal Vascular and Emergency Medicine-Stroke Unit, University of Perugia, Perugia, Italy

ARTICLE INFO

Keywords:

REMAP trials

Bayesian inference

Response adaptive randomization

ABSTRACT

Systems tend toward inertia until an external pressure pushes them toward change; thus, a situation of crisis such as the COVID-19 pandemic represents an opportunity for technological innovation. The prevailing need for treatments and vaccines has impelled innovation in the world of randomized clinical trials (RCT), resorting to ideas that had been floating around for a while. Is this merely a circumstantial phenomenon or are new methods here to stay?

1. Editorial

Systems tend toward inertia until an external pressure pushes them toward change; thus, a situation of crisis such as the COVID-19 pandemic represents an opportunity for technological innovation. The prevailing need for treatments and vaccines has impelled innovation in the world of randomized clinical trials (RCT), resorting to ideas that had been floating around for a while (Fig. 1). Is this merely a circumstantial phenomenon or are new methods here to stay?

The origin of the RCT dates back to the seminal work by Ronald Fisher regarding experimental designs, published in the 1920s, although the first RCT would not be conducted until 1948 [1]. The method has enabled medicine to progress since then. Nevertheless, RCTs pose several dilemmas, including being slow, relatively inefficient, complex and yielding broad conclusions (average treatment effects) that are scarcely generalizable to specific profiles [2]. This involves undeniable tension: while the clinician is urged to act on the basis of RCTs, clinical practice guidelines are plagued with recommendations based on inconclusive evidence [3]. The regulatory authorities are concerned about the type I error (the rejection of a true null hypothesis). At the bedside, the patient does not ask for irrefutable tests but for sensible decisions. For some researchers, traditional RCTs, because of these weaknesses and especially because of the delay in obtaining evidence,

are unsuitable for the tumultuous times of pandemic [3]. However, the need for evidence is always urgent when dealing with any seriously ill patient.

Faced with this dichotomy, the new philosophy is learning while doing, sacrificing certain praxis that are cast in stone for the sake of biomedical research [4]. Procedure quality must be safeguarded and auditable, regardless of the method, in line with the stringent regulations of good clinical practices, to ensure the quality of the data and all procedures.

This trend is embodied in the new designs of Randomized Embedded Multifactorial Adaptive Platform (REMAP) Trials (Table 1) [3]. The design is embedded because research and standard therapy merge, blurring the distinction between patients in clinical practice vs clinical trials, forasmuch as all require efficacious treatments and there is something to be learnt from every experience. The design is multifactorial because multiple treatment domains may be evaluated simultaneously, in parallel or sequentially, and adaptive inasmuch as it incorporates the observed response to modify treatment allocation probabilities (response adaptive randomization) [5]. The probability that the patient benefit from the clinical trial thereby increases, minimizing the number of individuals randomized to an inferior therapy, once this becomes clear by the end of the trial. On-the-fly trends from the RCT are used to adjust therapy assignment. The approach is

* Corresponding author at: Medical Oncology Department, Hospital Universitario Central de Asturias, Av. Roma, s/n, 33011 Oviedo, Asturias, Spain.

E-mail address: palucaji@hotmail.com (P. Jiménez-Fonseca).

<https://doi.org/10.1016/j.ejim.2021.11.002>

Received 12 October 2021; Accepted 6 November 2021

Available online 12 November 2021

0953-6205/© 2021 Published by Elsevier B.V. on behalf of European Federation of Internal Medicine.

mind-blowing because a smooth transition is made from RCT to reality.

The REMAP strategy generated following the COVID-19 crisis have been forged within multinational consortia financed with public funds, a network of networks of sites that research while simultaneously treating a multitude of patients. Unlike the classic RCT, REMAP trials are conceived as on-going, flexible platforms for learning about different therapies applied to different subgroups. The aim is to extrapolate the conclusions to strata of interest.

The reader accustomed to traditional RCTs is stuck by the highly adaptive design of REMAP trials, the lack of an explicit sample size, and the use of Bayesian statistics. Not surprisingly, most RCTs are analyzed with frequentist models, based on the analysis of the probability of the data under the model's hypotheses and assumptions. This inference is supported on assumptions that would arise and remain stable if the RCT were repeated a multitude of times (therein the term 'frequentist') [6]. Typically, a robust rationale must be constructed beforehand. The purpose is to probe the non-conformity of the data with respect to the null hypothesis, but not to analyze the direct support of the alternative hypothesis. Among the drawbacks of this frequentist approach is that to calculate statistical power, strict assumptions about the events and data structure must be made, complicating adaptations in the midterm trial if the pretreatment assumptions are wrong. This can result in an inconclusive RCT that limits learning, often mistaking the absence of evidence for the absence of effect [7].

Therefore, researchers must be mindful of the fact that the

Table 1

Main characteristics of clinical trial designs.

	Traditional RCT	REMAP
Randomization	Yes	Yes
Flexibility	Interim analyses	Adaptive
Sample size	Pre-specified	Variable
Treatment arms	Pre-defined	Variable
Analysis	Usually frequentist	Bayesian
Interpretation	Binary	Probabilistic
Integration into routine clinical practice	No	Yes. Embedded

frequentist approach, rigorously used in pivotal trials and drug registration, may not be the most suitable means by which to respond to research carried out during an on-going crisis. The most important lesson in the early stages of the Covid-19 pandemic was that clinicians were dealing with critically ill patients in highly uncertain conditions. The null hypothesis significance testing framework was a limited method in this situation, because any trial conducted hastily, almost heroically, had low statistical power and, as such, did not allow for learning from the data if the result was not statistically significant [8,9]. For example, with a p-value = 0.057, the investigator is basically obligated to discontinue judgement and no one gains anything positive from the experience [10].

Throughout the history of medicine, others have wisely argued that we should not settle for inconclusive results or hold back on sensible,



Fig. 1. Pathways of biomedical research.

urgent decisions because there is not a high level of evidence. For Bradford Hill, “All scientific work is incomplete—whether it be observational or experimental. All scientific work is liable to be upset or modified by advancing knowledge. That does not confer upon us a freedom to ignore the knowledge we already have, or to postpone the action that it appears to demand at a given time” [11].

To overcome these limitations, Bayesian statistics conform to REMAP trials, computing the probability of the hypotheses under the data available and the background information. This makes on-going learning possible, since the posterior probability distributions update naturally as the information accumulates. Bayesian methods emphasize the estimation of the true parameters of the hypotheses, allowing the actual probability of benefit or harm from therapy under conditions of uncertainty to be determined [7,12]. We can therefore estimate the probability of a drug working or not, so as to be able to make the most judicious decision, taking into account the full context. In fact, Bayesian models have helped us to navigate turbulent waters [13], and it is imperative that researchers be aware of their existence, particularly now [14]. Hence, during an epidemic of Ebola, researchers needed to acquire medical evidence incrementally; the timing was not compatible with rigid, often misinterpreted, decision rules that lead to all-or-nothing binary conclusions, therefore opting for Bayesian statistics [15].

These methods permit strict decision rules to be enacted. In the case of the community-acquired pneumonia REMAP-CAP trial, the decision rule was deemed to be that the intervention would be declared superior under a probability >0.99 of being the best therapy; inferior if the probability were <0.01 , and equivalent when 90% of the posterior distribution fell in the region of practical equivalence [3]. Notably, the sample size of an adaptive study is not fixed; it can be recalculated through simulations, although the platform can work perpetually until the required answers are obtained, according to these stopping rules or new questions emerge. This avoids inconclusive RCTs and enables information to be efficiently obtained regarding subgroup effects in all the studies [16].

As an example, an international collaboration (ATTACC, ACTIV-4a, and REMAP-CAP) have recently proven that anticoagulant therapy at therapeutic dosages is beneficial compared to thromboprophylaxis in non-critically ill patients with Covid-19, as an initial strategy in patients without thrombosis [17]. This trial demonstrated that the probability of benefit in terms of organ support-free days of therapeutic-dose anticoagulation was 98.6% (odds ratio, 1.27; 95% credible interval, 1.03–1.58).

The choice of suboptimal endpoints has been another of the most relevant features of this period. Hard endpoints, such as mortality, are often used in biomedical research to evaluate therapeutic effect. However, there are several circumstances in which the issue is less straightforward insofar as mortality does not capture the entire picture, in that it does not factor in the level of supportive care, including requirements for mechanical ventilation, ECMO, or ICU admission. The fundamental reason is that multiple scenarios, including infections, are clinically diverse and range between asymptomatic and serious events. Consequently, binary endpoints are occasionally incapable of revealing subtle, yet clinically relevant, effects of therapies [18], thereby associating remarkable loss of information [19]. Consequently, researchers must be aware that ordinal regression methods, frequentist or Bayesian, can benefit from much more efficient ordinal endpoints [20,21]. For instance, one recent randomized clinical trial (NCT01052480) compared the use of anti-influenza plasma plus standard care in patients with flu [22]. The trial’s primary endpoint was time to normalization of patients’ respiratory status, defined in binary form. This objective was not reached with a hazard ratio (HR) of 1.71 (95% CI 0.96–3.06) and, given the low frequency of death, no differences were seen in overall survival. Nevertheless, the authors observed that an ordinal scale would make it possible to establish the benefit of the experimental treatment with a decrease in progressively more severe outcomes compared to standard treatment. A family of models make it possible to evaluate

quantitative measurements, such as ordinal scales or interval scale variables. The most interpretable models assume that the effect of therapy is consistent across different levels of the endpoint, thus, the name “proportional odds models” [23,24]. Bayesian alternatives have been developed for these models.

The use of ordinal endpoints in COVID-19 research has been inconsistent in the first stages of the pandemic. More recently, Self et al. used a Bayesian ordinal model to assess the efficacy of hydroxychloroquine against placebo on clinical status [25]. Another trial with remdesivir did take the ordinal nature of clinical relief into account, but the changes were only evaluated at a single point in time [26]. Our group has successfully implemented Peterson & Harrell’s model to model quality of life in oncological patients [27].

The use of longitudinal models, which take variation over time into account, could enhance the efficiency of these RCTs. Accordingly, it is natural to think that ordinal Bayesian models could be useful in rational drug development during the COVID-19 pandemic, molding the rigor of the method to the need to take serious decisions on the spot as Bradford Hill claimed [11].

Will the XXI century be Bayesian as Lindley predicted [28]? Will we expand these flexible designs to other fields beyond? The challenge of precision medicine lies in attaining robust evidence in smaller and smaller strata, often clearly identified by biological platforms. To guarantee reasonable credibility in these subgroups, designs, methods, and platforms must be reformed [29].

The possible change in standard research methodology from frequentist RCTs to Bayesian REMAP trials must be carried out in an orderly fashion after resolving some of the problems associated with adaptive designs and Bayesian analyses. Nothing will prevent some RCTs from failing. For instance, on occasion, by coincidence or bad luck, the first steps of the study, when there are few patients admitted will lead us to improperly open or close treatment arms. The non-binary, probabilistic results of the Bayesian analysis will have to be made binary, whether we want to or not, establishing critical thresholds, although it will be wise to put off any judgment to the decision makers who will have to issue verdicts synthesizing probabilistic, context-based information. For some medical applications, the frequentist approach will continue to be the most fitting. It behooves us to integrate clinical trials into routine clinical practice, however, ultimately, including a patient in a clinical trial must be clearly circumscribed so that the entire research process aligns with Institutional Review Boards (IRBs) who must be convinced of this methodology after articulate explanation.

Finally, analogous to how one standard of care is replaced by another in the clinical setting, new designs, with positive and undoubtedly necessary facets, should be deemed a new standard of research once they have proven to be better than classic RCTs, without dismissing a period during which former standards coexist with the most recent ones, depending on the subject to be investigated and the question to be answered.

The technological momentum that supported the World War II effort consigned decisive advances to us in diverse fields such as nuclear fission, electronics, jet propulsion, radar, antibiotics, or vaccines. The legacy of the current crisis should serve to reform how we generate knowledge pragmatically in the face of serious processes, such as cancer or cardiovascular diseases. This could be the great intellectual revolution in medicine in the 21st century, comparable to the cultural leap made by Bradford Hill’s RCT in 1948.

Declaration of Competing Interest

None.

References

- [1] Chalmers I. Why the 1948 MRC trial of streptomycin used treatment allocation based on random numbers. *J R Soc Med* 2011;104:383–6. SAGE Publications Sage UK: London, England.
- [2] Angus DC. Fusing randomized trials with big data: the key to self-learning health care systems? *JAMA* 2015;314:767–8. American Medical Association.
- [3] Angus DC, Berry S, Lewis RJ, Al-Beidh F, Arabi Y, van Bentum-Puijk W, et al. The REMAP-CAP (randomized embedded multifactorial adaptive platform for community-acquired pneumonia) study. rationale and design. *Ann Am Thorac Soc* 2020;17:879–91. American Thoracic Society.
- [4] Angus DC. Optimizing the trade-off between learning and doing in a pandemic. *JAMA* 2020;323:1895–6. American Medical Association.
- [5] Hu F, Rosenberger WF. The theory of response-adaptive randomization in clinical trials. John Wiley & Sons; 2006.
- [6] Bland JM, Altman DG. Bayesians and frequentists. *Br Med J* 1998;317:1151–60.
- [7] Altman DG, Bland JM. Statistics notes: absence of evidence is not evidence of absence. *Bmj* 1995;311:485. British Medical Journal Publishing Group.
- [8] Cao B, Wang Y, Wen D, Liu W, Wang J, Fan G, Ruan L, Song B, Cai Y, Wei M, Li X, Xia J, Chen N, Xiang J, Yu T, Bai T, Xie X, Zhang L, Li C, Yuan Y, Chen H, Li H, Huang H, Tu S, Gong F, Liu Y, Wei Y, Dong C, Zhou F, Gu X, et al. A trial of Lopinavir–Ritonavir in adults hospitalized with severe Covid-19. *N Engl J Med* 2020. Massachusetts Medical Society.
- [9] Carmona-Bayonas A, Jimenez-Fonseca P, Castañón E. A trial of Lopinavir-Ritonavir in Covid-19. *N Engl J Med* 2020:382.
- [10] Kruschke J. Doing Bayesian data analysis: a tutorial with R, JAGS, and STAN. Academic Press; 2014.
- [11] Hill AB. The environment and disease: association or causation? Sage Publications; 1965.
- [12] Lewis RJ, Angus DC. Time for clinicians to embrace their inner Bayesian?: Reanalysis of results of a clinical trial of extracorporeal membrane oxygenation. *JAMA* 2018;320:2208–10. American Medical Association.
- [13] McGrayne SB. The theory that would not die: how Bayes' rule cracked the enigma code, hunted down Russian submarines, & emerged triumphant from two centuries of controversy. Yale University Press; 2011.
- [14] Zampieri FG, Damiani LP, Bakker J, Ospina-Tascón GA, Castro R, Cavalcanti AB, et al. Effects of a resuscitation strategy targeting peripheral perfusion status versus serum lactate levels among patients with septic shock. A Bayesian reanalysis of the ANDROMEDA-SHOCK trial. *Am J Respir Crit Care Med* 2020;201:423–9. American Thoracic Society.
- [15] Dodd LE, Proschan MA, Neuhaus J, Koopmeiners JS, Neaton J, Beigel JD, et al. Design of a randomized controlled trial for Ebola virus disease medical countermeasures: PREVAIL II, the Ebola MCM Study. *J Infect Dis* 2016;213:1906–13. Oxford University Press.
- [16] Castañón E, Sanchez-Arreaez A, Alvarez-Manceñido F, Jimenez-Fonseca P, Carmona-Bayonas A. Critical reappraisal of phase III trials with immune checkpoint inhibitors in non-proportional hazards settings. *Eur J Cancer* 2020;136:159–68. Elsevier.
- [17] ATTACC and REMAP-CAP Investigators A-4a. Therapeutic anticoagulation with heparin in noncritically ill patients with Covid-19. *N Engl J Med* 2021;385:790–802. Mass Medical Soc.
- [18] Su S, Chaves SS, Perez A, D'Mello T, Kirley PD, Yousey-Hindes K, et al. Comparing clinical characteristics between hospitalized adults with laboratory-confirmed influenza A and B virus infection. *Clin Infect Dis* 2014;59:252–5. Oxford University Press.
- [19] Deysi BA, Kosinski AS, Snapinn SM. Power considerations when a continuous outcome variable is dichotomized. *J Biopharm Stat* 1998;8:337–52. Taylor & Francis.
- [20] Roozenbeek B, Lingsma HF, Perel P, Edwards P, Roberts I, Murray GD, et al. The added value of ordinal analysis in clinical trials: an example in traumatic brain injury. *Crit Care* 2011;15:R127. Springer.
- [21] Horby P, Wang Y, Fan G, Hayden F, Li Q, Wu Q, et al. Comparative outcomes of adults hospitalized with seasonal influenza A or B Virus Infection: application of the 7-category ordinal scale. *Open Forum Infect Dis* 2019. Oxford University Press.
- [22] Beigel JH, Tebas P, Elie-Turenne M-C, Bajwa E, Bell TE, Cairns CB, et al. Immune plasma for the treatment of severe influenza: an open-label, multicentre, phase 2 randomised study. *Lancet Respir Med* 2017;5:500–11. Elsevier.
- [23] Harrell FE. Ordinal logistic regression. *Regres Model Strateg* 2015:311–25.
- [24] Walker SH, Duncan DB. Estimation of the probability of an event as a function of several independent variables. *Biometrika* 1967;54:167–79. Oxford University Press.
- [25] Casey JD, Johnson NJ, Semler MW, Collins SP, Aggarwal NR, Brower RG, et al. Rationale and design of ORCHID: a randomized placebo-controlled clinical trial of hydroxychloroquine for adults hospitalized with COVID-19. *Ann Am Thorac Soc* 2020;17:1144–53. American Thoracic Society.
- [26] Beigel JH, Tomashek KM, Dodd LE, Mehta AK, Zingman BS, Kalil AC, Hohmann E, Chu HY, Luetkemeyer A, Kline S, Lopez de Castilla D, Finberg RW, Dierberg K, Tapson V, Hsieh L, Patterson TF, Paredes R, Sweeney DA, Short WR, Touloumi G, Lye DC, Ohmagari N, Oh M, Ruiz-Palacios GM, Benfield T, Fätkenheuer G, Kortepeter MG, Atmar RL, Creech CB, Lundgren J, et al. Remdesivir for the treatment of Covid-19 — preliminary report. *N Engl J Med* 2020. <https://doi.org/10.1056/NEJMoa2007764> [Internet] Massachusetts Medical Society Available from.
- [27] Carmona-Bayonas A, Calderón C, Hernández R, Montes AF, Castelo B, Ciria-Suarez L, Antónanzas M, Rogado J, Pacheco-Barcia V, Martínez EA. Prediction of quality of life in early breast cancer upon completion of adjuvant chemotherapy. *NPJ Breast Cancer* 2021;7:1–9. Nature Publishing Group.
- [28] Lindley DV. The future of statistics: a Bayesian 21st century. *Adv Appl Probab* 1975;7:106–15. JSTOR.
- [29] Burke JF, Sussman JB, Kent DM, Hayward RA. Three simple rules to ensure reasonably credible subgroup analyses. *Bmj* 2015:351. British Medical Journal Publishing Group.