CORRESPONDENCE

Randomization versus Real-World Evidence

TO THE EDITOR: Collins et al. (Feb. 13 issue)[1] set up a false dichotomy in presenting real-world evidence (RWE) as an "unreliable" substitute for findings from randomized, controlled trials (RCTs). While the "magic" of randomization ensures balance between groups in RCTs, it cannot ensure that outcomes are representative of a given population. In the United States, RCTs involving patients with cancer represent less than 5% of U.S. adults with cancer. Patients in RCTs are also younger, healthier, and less diverse than the other 95% of patients with cancer.[2] Because it is not always feasible to conduct an RCT, particularly among patients with rare cancers or genomic subtypes, the Food and Drug Administration (FDA) has accepted RWE from expanded-access studies, medical records, and insurance claims.[3,4] The 21st Century Cures Act, approved by Congress in 2016, which calls for the FDA to evaluate the potential of RWE in reviewing and approving new indications for existing drugs and in fulfilling postapproval study requirements, recognizes the value of RWE beyond its established role in postmarketing surveillance.[5] RWE offers insight into the ways in which a drug is actually used, barriers to use, and outcomes regarding toxicity and efficacy in daily medical practice. RWE both supplements and complements information gleaned from the prevailing standard of RCTs; the myth is that we must choose one option or the other.

Ajeet Gajra, M.D.
Marjorie E. Zettler, Ph.D., M.P.H.
Bruce A. Feinberg, D.O.
Cardinal Health Specialty Solutions
Dublin, OH
ajeet.gajra@cardinalhealth.com

Dr. Gajra reports being employed by Cardinal Health and ICON; and Drs. Zettler and Feinberg, being employed by Cardinal Health. No other potential conflict of interest relevant to this letter was reported.

This letter was published on July 23, 2020, at NEJM.org.

1. Collins R, Bowman L, Landray M, Peto R. The magic of randomization versus the myth of real-world evidence. N Engl J Med 2020;382:674-8.
2. Unger JM, Cook E, Tai E, Bleyer A. The role of clinical trial participation in cancer research: barriers, evidence, and strategies. Am Soc Clin Oncol Educ Book 2016;35:185-98.
3. Food and Drug Administration. Lutetium Lu 177 dotatate (Lutathera) drug approval package. 2018 (https://www.accessdata.fda.gov/drugsatfda_docs/nda/2018/208700Orig1s000MultidisciplineR.pdf).
4. Food and Drug Administration. Palbociclib S-008 (Ibrance) drug approval package. 2019 (https://www.accessdata.fda.gov/drugsatfda_docs/nda/2019/207103Orig1s008.pdf).
5. Public Law 114-255, 114th Congress. The 21st Century Cures Act. December 13, 2016 (https://www.congress.gov/114/plaws/publ255/PLAW-114publ255.pdf).

DOI: 10.1056/NEJMc2020020

TO THE EDITOR: Collins et al. make a strong plea for randomization, yet it is not strong enough: a twofold difference in effects is mentioned as a threshold beyond which randomization would not be needed. In the recent past, some observational studies have reported an increase in mortality with digoxin in patients with heart failure, atrial fibrillation, or both. Figure 1 shows that the results of the observational studies listed were so conflicting that it would be impossible to identify a true effect. Even worse, bias by indication can be found in the only RCT identified in our search,[2] in which patients who received pretreatment with digoxin were compared with those had not.[3] Notably, the effect of digoxin on mortality was neutral. There is a second message in Figure 1: all authors were convinced that they had properly adjusted for population differences, yet the overall effect of adjustment was rather small. The magnitude of an effect estimate may be a useful criterion for screening purposes when pharmacovigilance is practiced,[4,5] but statisticians need to lower their expectations regarding what can be achieved

Author	Journal and Year	Indication	Comments	Unadjusted Estimate	Adjusted Estimate
Banerjee	Int J Cardiol 1996	HF		6.40 (0.00–36.00)	
Moss	Circulation 1981	MI	*	4.35 (2.41–7.86)	
Al-Zakwani	J Cardiovasc Pharmacol Ther 2016	AF	Odds ratio		4.22 (2.23–7.99)
Ryan	Circulation 1983	Other	CAD*	3.60 (3.22–4.03)	
Oster	J Am Heart Assoc 2016	Other	Infants, congenital heart disease		3.50 (1.10–11.70)
Leor	Cardiovasc Drugs Ther 1995	MI			2.80 (1.80–4.20)
Muller	N Engl J Med 1986	MI	*	2.58 (1.92–3.46)	
Byington	J Am Coll Cardiol 1985	MI	*	2.49 (1.86–3.33)	
Moss	Cardiovasc Drugs Ther 1987	MI			2.30 (1.40–3.70)
Pastori	Int J Cardiol 2015	AF			2.22 (1.42–3.48)
Madsen	J Am Coll Cardiol 1984	MI	*	2.21 (1.74–2.81)	
Lindsay	Lancet 1999	HF			2.12 (1.21–3.74)
Rivinius	Ther Clin Risk Manag 2017	Other	Digoxin before heart transplantation		2.10 (1.04–4.25)
Køber	Eur Heart J 1994	MI			1.80 (1.20–2.50)
Al-khateeb	Int J Cardiol 2017	HF		1.81 (1.33–2.45)	1.74 (1.20–2.38)
Georgiopoulou	Circ Heart Fail 2009	HF	Death or LVAD implantation	2.28 (1.51–3.43)	1.73 (1.09–2.75)
Freeman	Circ Cardiovasc Qual Outcomes 2013	HF			1.72 (1.25–2.36)
Erath	Eur Heart J Cardiovasc Pharmacother 2016	Other	ICD recipients	2.47 (1.87–3.25)	1.65 (1.14–2.39)
Freeman	Circ Arrhythm Electrophysiol 2015	AF		1.71 (1.52–1.93)	1.62 (1.43–1.84)
Diller	Eur Heart J 2016	Other	Eisenmenger's syndrome	1.59 (0.85–2.98)	
Bigger	Am J Cardiol 1985	MI			1.58 (0.97–2.60)
Adedinsewo	Circ Arrhythm Electrophysiol 2017	AF		1.68 (1.57–1.80)	1.54 (1.39–1.70)
Coiro	Clin Res Cardiol 2017	MI		2.24 (2.05–2.44)	1.54 (1.42–1.67)
Gjesdal	Heart 2008	AF		1.58 (1.29–1.94)	1.53 (1.22–1.92)
Pasca	J Am Soc Echocardiogr 2016	Other	Degenerative mitral stenosis		1.51 (1.22–2.04)
Domanski	J Card Fail 2005	HF	Women		1.48 (1.09–2.00)
Hallberg	Eur J Clin Pharmacol 2007	AF		1.83 (1.67–2.00)	1.42 (1.29–1.56)
Rodríguez-Mañero	Rev Esp Cardiol (Engl Ed) 2014	AF			1.42 (0.77–2.60)
Mølstad	Eur Heart J 1991	MI		2.33 (1.80–3.00)	1.41 (1.07–1.86)
Whitbeck	Eur Heart J 2013	AF			1.41 (1.19–1.64)
Al-Zakwani	J Cardiovasc Pharmacol Ther 2016	AF+HF	Odds ratio		1.37 (0.74–2.53)
Ghali	J Card Fail 2004	HF	*	1.37 (1.04–1.81)	
Domanski	J Card Fail 2005	HF	Men		1.37 (1.20–1.56)
Reicher-Reiss	Int J Cardiol 1999	MI	MI survivors with CAD*	2.71 (2.35–3.27)	1.33 (0.97–1.82)
Casiglia	Jpn Heart J 1998	AF	*	1.31 (0.55–3.14)	
Butler	Congest Heart Fail 2010	HF		1.46 (1.23–1.64)	1.28 (1.05–1.57)
Casiglia	Jpn Heart J 1998	Other	Neither HF nor AF*	1.45 (1.32–1.60)	1.27 (1.07–1.48)
Turakhia	J Am Coll Cardiol 2014	AF		1.37 (1.33–1.40)	1.26 (1.23–1.29)
Flahavan	BJU Int 2014	Other	Prostate cancer	2.11 (1.86–2.41)	1.24 (1.07–1.43)
Eisen	J Am Heart Assoc 2017	AF			1.22 (1.12–1.34)

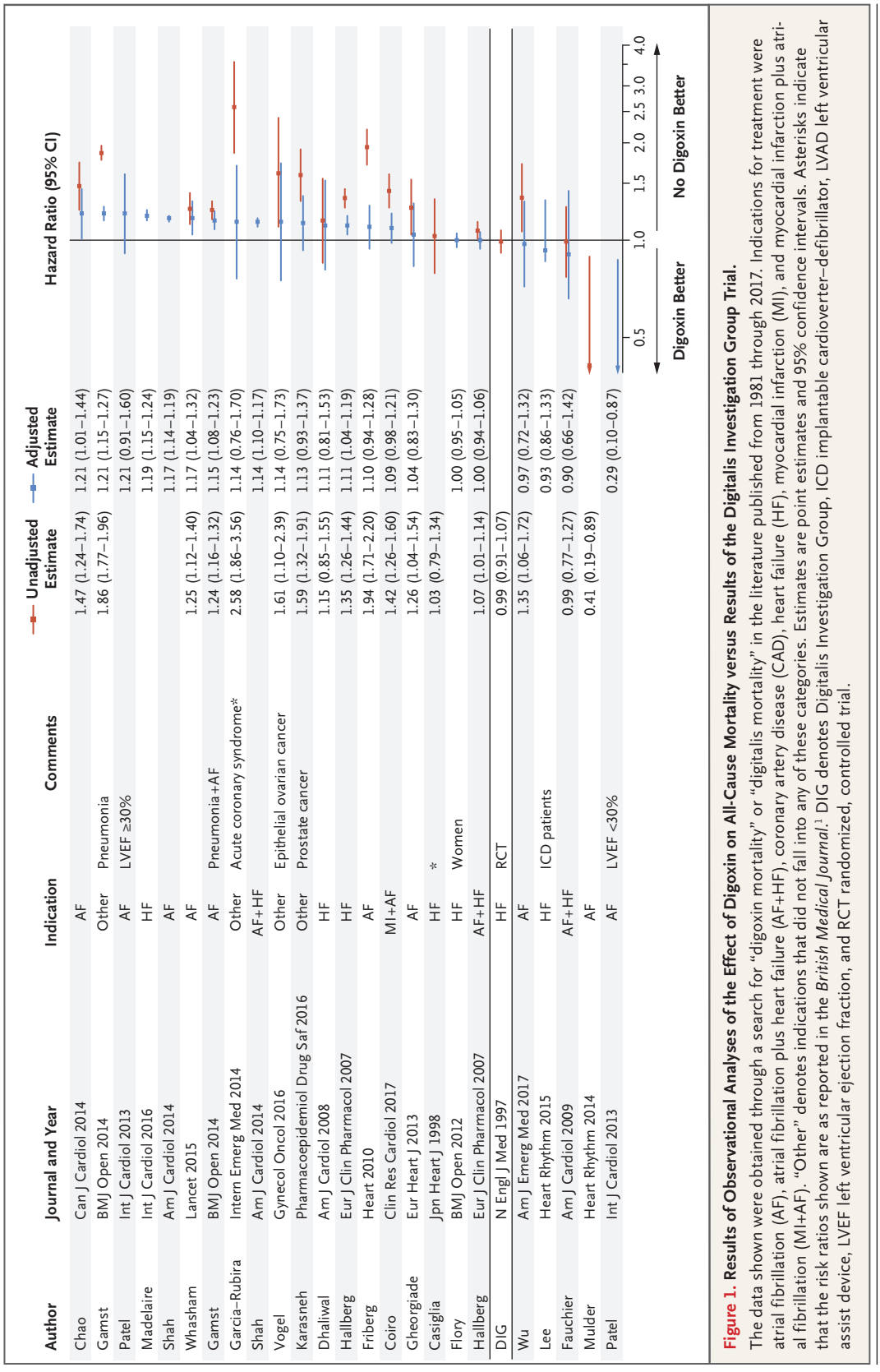

Figure 1. Results of Observational Analyses of the Effect of Digoxin on All-Cause Mortality versus Results of the Digitalis Investigation Group Trial.

The data shown were obtained through a search for "digoxin mortality" or "digitalis mortality" in the literature published from 1981 through 2017. Indications for treatment were atrial fibrillation (AF), atrial fibrillation plus heart failure (AF+HF), coronary artery disease (CAD), heart failure (HF), myocardial infarction (MI), and myocardial infarction plus atrial fibrillation (MI+AF). "Other" denotes indications that did not fall into any of these categories. Estimates are point estimates and 95% confidence intervals. Asterisks indicate that the risk ratios shown are as reported in the *British Medical Journal*.[1] DIG denotes Digitalis Investigation Group, ICD implantable cardioverter–defibrillator, LVAD left ventricular assist device, LVEF left ventricular ejection fraction, and RCT randomized, controlled trial.

when statistical adjustment is used to account for population differences.

Lukas Aguirre Dávila, Dr. rer. nat.

Paul-Ehrlich-Institut
Langen, Germany
aguirredavila.lukas@gmx.de

Armin Koch, Dr. sc. hum.
Florian Lasch, Dr. rer. nat.

Hannover Medical School
Hannover, Germany

Drs. Aguirre Dávila and Koch report being responsible for the biostatistical planning and design of the Digitoxin to Improve Outcomes in Patients with Advanced Chronic Heart Failure (DIGIT-HF) trial (Eudra-CT number, 2013-005326-38); and Dr. Koch, being a member of the steering committee. The trial was funded by the Federal Ministry of Education and Research in Germany. No other potential conflict of interest relevant to this letter were reported.

This letter was published on July 23, 2020, at NEJM.org.

1. Ziff OJ, Lane DA, Samra S, et al. Safety and efficacy of digoxin: systematic review and meta-analysis of observational and controlled trial data. BMJ 2015;351:h4451.
2. The Digitalis Investigation Group. The effect of digoxin on mortality and morbidity in patients with heart failure. N Engl J Med 1997;336:525-33.
3. Aguirre Dávila L, Weber K, Bavendiek U, et al. Digoxin-mortality: randomized vs. observational comparison in the DIG trial. Eur Heart J 2019;40:3336-41.
4. Ventola CL. Big data and pharmacovigilance: data mining for adverse drug events and interactions. P T 2018;43:340-51.
5. European Medicines Agency. Guideline on good pharmacovigilance practices (GVP) Module IX Addendum I: methodological aspects of signal detection from spontaneous reports of suspected adverse reactions (EMA/209012/2015). October 9, 2017 (https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-good-pharmacovigilance-practices-gvp-module-ix-addendum-i-methodological-aspects-signal_en.pdf).

DOI: 10.1056/NEJMc2020020

TO THE EDITOR: Collins et al. suggest that giving researchers access to electronic health records systems and other, specialized registries for the purpose of identifying potential trial participants would facilitate the conduct of RCTs. The conduct of low-risk, pragmatic RCTs should be considered to assess the comparative effectiveness of marketed medicines. In these trials, usual clinical practices would be observed, with no or minimal increase in risk to patients as compared with that associated with normal clinical practice.[1] Trial conduct would be facilitated if the requirement to obtain written informed consent were modified or removed from European legislation when certain requirements are met, an approach that is already being considered in the United States[2] and by the Council for International Organizations of Medical Sciences.[3] The authors rather lightly dismiss the problem of generalizing the results of RCTs to broader populations, stating that proportional treatment effects can be generalized "unless there is good reason to expect otherwise." However, generalizability may become a significant challenge when trial participants do not resemble the typical patient in routine clinical practice.[4] Low-risk, pragmatic RCTs with modified or waived consent may help to combine the magic of randomization with the myth of RWE.

Rafael Dal-Ré, M.D., Ph.D.

Health Research Institute–Fundación Jiménez Díaz University Hospital
Madrid, Spain
rafael.dalre@quironsalud.es

Robert J. Mentz, M.D.

Duke Clinical Research Institute
Durham, NC

Frits R. Rosendaal, M.D., Ph.D.

Leiden University Medical Center
Leiden, the Netherlands

Dr. Dal-Ré reports receiving lecture fees from Palex Medical; and Dr. Mentz, receiving grants and consulting fees from Novartis, Amgen, AstraZeneca, Merck, and Bayer, consulting fees from Sanofi, Abbott, and Boehringer Ingelheim, and grants from American Regent and Medtronic. No other potential conflict of interest relevant to this letter was reported.

This letter was published on July 23, 2020, at NEJM.org.

1. Dal-Ré R, Avendaño-Solà C, Bloechl-Daum B, et al. Low risk pragmatic trials do not always require participants' informed consent. BMJ 2019;364:l1092.
2. Food and Drug Administration Office of Good Clinical Practice. IRB waiver or alteration of informed consent for clinical investigations involving no more than minimal risk to human subjects: guidance for sponsors, investigators and institutional review boards. July 2017 (https://www.fda.gov/media/106587/download).
3. Council for International Organizations of Medical Sciences (CIOMS). International ethical guidelines for health-related research involving humans. 2016 (https://cioms.ch/shop/product/international-ethical-guidelines-for-health-related-research-involving-humans/).
4. Levi M, Hovingh GK, Cannegieter SC, Vermeulen M, Büller HR, Rosendaal FR. Bleeding in patients receiving vitamin K antagonists who would have been excluded from trials on which the indication for anticoagulation was based. Blood 2008;111:4471-6.

DOI: 10.1056/NEJMc2020020

TO THE EDITOR: Collins et al. highlight the indisputable need to reduce the complexities of RCTs and caution against adopting RWE from nonrandomized studies as an alternative to RCTs. The International Society for Pharmacoepidemiology agrees with both points but views RCTs and RWE

as complementary, not magic and myth but yin and yang, and aligned on a continuum of imperfect approaches in scientific inquiry.

RWE is useful for the study of unintended drug effects[1] and can supplement the data provided in RCTs regarding effectiveness. Real-world data facilitate RCT recruitment and follow-up and reflect routine medical practices. RWE may be the only option when recruitment, randomization, or preservation of randomization is infeasible. RWE methods can be used to evaluate variation in treatment effects and to accommodate deviations from RCT protocols, whereas the stringent controls in RCTs intended to prevent bias come with a risk of inductive fallacies.

If properly designed in the same target population, RWE and RCTs findings align,[2-4] but as with RCTs, RWE must be grounded in the scientific method. As such, RCTs and RWE constitute interwoven sciences, with the shared aim of helping regulators and clinicians understand the benefits and risks of medical interventions.

Almut G. Winterstein, Ph.D.

University of Florida
Gainesville, FL
almut@ufl.edu

Cynthia J. Girman, Dr.P.H.

CERobs Consulting
Chapel Hill, NC

Anton Pottegård, Ph.D.

Syddansk Universitet
Odense, Denmark

Dr. Winterstein reports being President of the International Society of Pharmacoepidemiology (ISPE); Dr. Girman, serving on the ISPE board of directors and as cochair of the ISPE work group on RWE and regulatory decisions; and Dr. Pottegård, serving as cochair of the ISPE work group on RWE reproducibility and transparency. No potential conflict of interest relevant to this letter was reported.

This letter was published on July 23, 2020, at NEJM.org.

1. Vandenbroucke JP. When are observational studies as credible as randomised trials? Lancet 2004;363:1728-31.
2. Hernán MA, Robins JM. Per-protocol analyses of pragmatic trials. N Engl J Med 2017;377:1391-8.
3. Lund JL, Richardson DB, Stürmer T. The active comparator, new user study design in pharmacoepidemiology: historical foundations and contemporary application. Curr Epidemiol Rep 2015;2:221-8.
4. Franklin JM, Glynn RJ, Martin D, Schneeweiss S. Evaluating the use of nonrandomized real-world data analyses for regulatory decision making. Clin Pharmacol Ther 2019;105:867-77.

DOI: 10.1056/NEJMc2020020

TO THE EDITOR: Collins et al. highlight the need to improve the conduct and feasibility of randomized trials. Two points deserve discussion. First, we agree that randomization is "guaranteed to result in groups of patients that are balanced (give or take the play of chance)." However, "the play of chance" is not trivial. If the prevalence of each of five prognostic factors in a 400-person trial is 20%, then there is a 23% chance that there will be a between-group imbalance of more than 10% for at least one covariate.[1] Coefficients from regression models that adjust for baseline differences do not represent the causal effect of treatment if the treatment effect is different between patients with different prognostic factors.[2] Second, if adherence is less than 100%, then an intention-to-treat analysis from a trial that assesses the effect of assigning treatment will underestimate the effect of receiving treatment. Estimating the effect of receiving treatment in a trial requires treating the trial data as an observational study, with all the necessary assumptions. A well-conducted observational study that directly assesses the effect of receiving treatment may be more accurate.[3,4]

Ian Shrier, M.D., Ph.D.

McGill University
Montreal, QC, Canada
ian.shrier@mcgill.ca

Steven D. Stovitz, M.D.

University of Minnesota
Minneapolis, MN

No potential conflict of interest relevant to this letter was reported.

This letter was published on July 23, 2020, at NEJM.org.

1. Shrier I, Boivin JF, Steele RJ, et al. Should meta-analyses of interventions include observational studies in addition to randomized controlled trials? A critical examination of underlying principles. Am J Epidemiol 2007;166:1203-9.
2. Shrier I, Redelmeier A, Schnitzer M, Steele RJ. Challenges in interpreting results from 'multiple regression' when there is interaction between covariates. BMJ Evid Based Med 2019 August 22 (Epub ahead of print).
3. Hernán MA, Robins JM. Per-protocol analyses of pragmatic trials. N Engl J Med 2017;377:1391-8.
4. Shrier I, Steele RJ, Verhagen E, Herbert R, Riddell CA, Kaufman JS. Beyond intention to treat: what is the right question? Clin Trials 2014;11:28-37.

DOI: 10.1056/NEJMc2020020

TO THE EDITOR: We agree with Collins et al. that RCTs are the standard for the estimation of treat-

ment effects and that improvements are needed. However, referring to observational studies as untrustworthy does not do justice to more than 150 years of epidemiologic study, starting with John Snow's study of water pumps and cholera. Formal frameworks for causal inference provide systematic approaches for addressing potential biases (e.g. confounding and selection) and for evaluating the assumptions required to estimate treatment effects, regardless of their magnitude.[1] Failure to follow these frameworks can result in biased estimates and misleading conclusions in both RCTs and observational studies. Application of these methods to 733,804 electronic health records suggested that statins did not protect against cancer, a finding consistent with that of prior RCTs.[2] Randomization is often unethical or impractical in public health and medicine. Furthermore, RCTs are not without challenges (e.g., unpreventable and differential missingness), and generalizability is more complicated than simply assuming proportional effects.[3] Sensitivity analyses require more than an examination of the degree of changes in incidence.[4] The way to address these challenges lies not in choosing exclusively observational studies or RCTs but in building consensus across different study designs and in the application of rigorous methods that allow us to extract the truth from the data.

Laura B. Balzer, Ph.D.

University of Massachusetts–Amherst
Amherst, MA
lbalzer@umass.edu

Francesca Dominici, Ph.D.

Harvard T.H. Chan School of Public Health
Boston, MA

No potential conflict of interest relevant to this letter was reported.

This letter was published on July 23, 2020, at NEJM.org.

1. Petersen ML, van der Laan MJ. Causal models and learning from data: integrating causal modeling and statistical estimation. Epidemiology 2014;25:418-26.
2. Dickerman BA, García-Albéniz X, Logan RW, Denaxas S, Hernán MA. Avoidable flaws in observational analyses: an application to statins and cancer. Nat Med 2019;25:1601-6.
3. Bareinboim E, Pearl J. A general algorithm for deciding transportability of experimental results. J Causal Inference 2013;1:107-34.
4. Imbens GW. Sensitivity to exogeneity assumptions in program evaluation. Am Econ Rev 2003;93:126-32.

DOI: 10.1056/NEJMc2020020

THE AUTHORS REPLY: We agree that there are circumstances in which it is not feasible to evaluate the effects of treatment in randomized trials. However, even in such circumstances, there is a need for greater recognition — by regulators and policy makers as well as by scientists and clinicians — that dependence on observational studies risks inadequate assessment of not only the efficacy of treatment but also its safety owing to the potential biases inherent in such studies.

For example, on the basis of information provided in observational studies, the FDA authorized hydroxychloroquine for the treatment of Covid-19.[1] Subsequently, on the basis of adverse trends reported in other such studies (some of which were found to be flawed in additional ways), regulatory authorities and commentators[2,3] put pressure on investigators to stop large randomized trials designed to provide reliable assessment of the safety and efficacy of hydroxychloroquine. Fortunately, despite this undue reliance on observational evidence, sufficient data had already accrued in the randomized trials to convincingly indicate the lack of benefit.[4]

The problem is not that observational studies cannot get the right answer when treatment effects are null or only moderate, it is that they cannot be relied on to have done so. Moreover, although observational studies cannot generally be trusted when there is less than a twofold difference in the incidence of a health outcome between people who have had a particular treatment and those who have not, this does not necessarily mean that more extreme relative risks in such studies can be trusted (although, as stated in our article, large effects on rare outcomes may be more likely to be causal).

We agree that observational data can be used to facilitate the recruitment of patients into randomized trials and to enhance follow-up during and after the scheduled treatment period. In addition, such data can help in generalizing the results of randomized trials to different patient populations. However, instead of promoting the replacement of randomized trials with unreliable observational treatment effects, what is now needed is both the removal of unnecessary bureaucratic obstacles to randomized trials that greatly increase cost and complexity and the implementa-

tion of innovative strategies that facilitate the conduct of better randomized trials.

As we discussed, regulatory strategies should be updated to make the recruitment of a wide range of patients into randomized trials a more rapid, straightforward process. Such an initiative would yield reliable and generalizable evidence regarding treatment safety and efficacy on which doctors and patients can rely.

Rory Collins, F.R.S.
Louise Bowman, M.D., F.R.C.P.
Martin Landray, Ph.D., F.R.C.P.

University of Oxford
Oxford, United Kingdom
rory.collins@ndph.ox.ac.uk

Since publication of their article, the authors report no further potential conflict of interest.

This letter was published on July 23, 2020, at NEJM.org.

1. Hinton DM, Food and Drug Administration. Letter to Dr. Rick Bright, Director, Biomedical Advanced Research and Development Authority re: FDA emergency use authorization (EUA) of chloroquine and hydroxychloroquine. March 28, 2020 (revoked) (https://www.fda.gov/media/136534/download).
2. COVID-19: reminder of the risks of chloroquine and hydroxychloroquine. Press release of the European Medicines Agency, May 29, 2020 (https://www.ema.europa.eu/en/news/covid-19-reminder-risks-chloroquine-hydroxychloroquine).
3. Topol E. Comments on hydroxychloroquine. May 22, 2020 (https://twitter.com/EricTopol/status/1263820223443898369?s=20).
4. Statement from the Chief Investigators of the Randomised Evaluation of COVid-19 thERapY (RECOVERY) Trial on hydroxychloroquine: no clinical benefit from use of hydroxychloroquine in hospitalised patients with COVID-19. June 5, 2020 (https://www.recoverytrial.net/files/hcq-recovery-statement-050620-final-002.pdf).

DOI: 10.1056/NEJMc2020020