Randomized Clinical Trial or Real-World Evidence: How Historical Events, Public Demand, and the Resulting Laws and Regulations Shaped the Body of Medical Evidence

John I. Mackowiak, PhD,[a] Christina D. Mack, PhD, MSPH,[b] Debra E. Irwin, PhD, MSPH,[c] and Robert Zura, MD[d]

Summary: The signing of the 21st Centuries Cures Act in 2016 was a confirmational step in a long journey toward an understood use and need for real-world evidence (RWE), even though the Food and Drug Administration (FDA) had the legislative authority to accept RWE since 1962 to demonstrate efficacy. The 21st Century Cures Act, as well as the subsequent FDA guidance published in 2017 and other supporting guidance, documents that since are opening the doors for the clinical and research community. They specifically allow for labeling changes and indication expansion based on RWE. The legislative discussion of efficacy requirements started in the late 1950s, when evidence of effectiveness was not required in the United States before the marketing of a drug or medical device, and calls for the real-world comparative effectiveness research were being made by Senator Estes Kefauver. When the thalidomide tragedy stuck, Congress and the Kennedy Administration rushed to pass a new law to require that drugs be "effective in use." The regulations subsequently drafted by the FDA to enforce the law often required placebo-controlled, randomized clinical trials (RCTs). In the 1980s, some started to label the RCT as the gold standard for medical evidence. The use of real-world data for new indication approval was not specifically prohibited by the 1962 law, but the new 2016 law sent a clear mandate to FDA, requiring the agency to review new forms of evidence such as RWE.

Key Words: real-world evidence, randomized clinical trial, 21st centuries cures act

(*J Orthop Trauma* 2021;35:S17–S21)

Accepted for publication December 9, 2020.

From the [a]Center for Outcomes Research, Cedar Point, NC; [b]IQVIA Real-World Solutions, Durham, NC; [c]IBM Watson Health, Durham, NC; and [d]Department of Orthopaedics, Louisiana State University Health Sciences Center, New Orleans, LA.

J. Mackowiak is a research consultant to Bioventus. C. D. Mack is a full-time employee of IQVIA, which is a research organization contracted with Bioventus on the BONES study. She holds stock in Johnson & Johnson and AstraZeneca. D. E. Irwin is an employee of IBM Watson Health and IBM Watson Health received financial support from Bioventus to conduct the BONES study. R. Zura is a consultant for Bioventus, Stryker and Osteocentric. He is a Board Member for the Kuntscher Society and the Louisiana Orthopaedic Association and serves on the Editorial Board of Orthopedics.

Reprints: John Mackowiack, PhD, Center for Outcomes Research, 186 Bluff Rd, Cedar Point, NC 28584 (e-mail: jmackowiak@c-o-r.org).

DOI: 10.1097/BOT.0000000000002040

1962 LAW REQUIRES DRUGS BE "EFFECTIVE IN USE"

Senator Estes Kefauver of Madisonville Tennessee, son of a struggling small-town hardware merchant and hotel owner, was holding 17 months of hearings in the Senate Antitrust and Monopoly Subcommittee during the late 1950s to address high drug prices and dubious claims of efficacy. His proposed legislation would prevent marketing of me-too drugs by requiring all newly approved drugs to have a comparative efficacy advantage over the existing similar molecule. He did not have the Kennedy administration's support, so the bill never saw the light of day as initially proposed.[1] Kennedy's lack of support for Kefauver's bill in the late 50s might have been colored by the fact that Kefauver narrowly beat Kennedy on the second ballot in 1956 (before presidential primaries) when Kennedy wanted to be the vice president running mate to Stevenson for President. But in an unpredicted turn of events, there suddenly was a political desire to respond to the thalidomide disaster. Kefauver's failed bill was resurrected and modified significantly to only require efficacy over placebo. The American Medical Association opposed the regulation of efficacy by a government body, stating that the "only possible final determination as to the efficacy and ultimate use of a drug is the extensive clinical use of that drug by large numbers of the medical profession over a long period of time."[1] Nevertheless, the law was passed unanimously in the Senate and Congress and was signed into law by Kennedy 8 days later.[1] Before that law was passed in 1962, only safety had to be demonstrated to allow marketing. The exact wording of that law which is still in effect today states:[2]

> SEC. 505. [21 U.S.C. 355]. No person shall introduce or deliver for introduction into interstate commerce any new drug, unless an approval of an application … is effective with respect to such drug. Any person may file with the Secretary an application with respect to any drug subject to the provisions of subsection. Such persons shall submit to the Secretary as a part of the application, full reports of investigations which have been made to show whether or not such drug is safe for use and whether such drug is effective in use; ….

The wording of the law "effective in use" could have been interpreted to mean effective in real-world use, but the

regulations that were written then and those in place today do not yet place an emphasis on real-world use. Efficacy, as defined by Archie Cochrane, was and is emphasized over effectiveness.[3]

Food and Drug Administration Regulations Required Controlled Clinical Trials

Dr Louis Lasagna, MD, the father of clinical pharmacology,[4] and others, testified before congress in favor of controlled clinical trials for proving a drug's effectiveness proof.[5] The Food and Drug Administration (FDA) was charged with writing the regulations that would be used to enforce the new law. They relied on many experts, but Dr Louis Lasagna, author of the commonly used 1962 text book "Clinical Pharmacology", from nearby Johns Hopkins, was thought to highly influence the FDA to favor placebo-controlled, randomized, control trials as a means to demonstrate safety and efficacy. His 1954 article entitled, "A Study of the placebo response" was ranked by Richard Horton, editor of Lancet, as 1 of the 27 most notable achievements in a medical canon going back to Hippocrates. From his training as an anesthesiologist, his early research, he knew how short-term pain can often be treated with placebo. This may have influenced his preference for placebo-controlled studies to demonstrate efficacy, over real-world studies to document if a treatment is "effective in use." The regulations required the phase I, II, III trial system that is in place today. Soon other countries adopted a similar system. This system of phased studies has worked well for diseases we were able to treat at that time. It took the medical community from a place where there was very little efficacy data, to a place where the commonly used small molecules and common devices had significantly more scientific data to support use, especially in comparison to placebo, in the population that was selected for study. But it did not provide evidence of effectiveness for all the diseases or indications where the product can be used, nor did it provide data for all population subgroups where the product could be used.

The RCT has a relatively short history. The first results of a RCT were published in 1948, a study of streptomycin for pulmonary fibrosis.[6] Thirty-four years later in 1982 was the first time the RCT was mentioned in medical literature as the gold standard; however, the authors in that article also state, "epidemiological research has become increasingly important because it offers a substitute for the unattainable scientific gold standard of a randomized experimental trail."[7] In 1979, Bonchek encouraged surgeons in an article regarding coronary bypass "to resist the almost religious fervor of those who would sanctify RCTs as the only means of learning the truth."[8]

The regulations to enforce section 505 of the 1962 law are listed in CFR 314.126.[9] They have evolved over time and may continue to evolve. They currently require an "adequate and well-controlled" study to demonstrate effectiveness. Although the regulations allow for the control to be (1) *placebo concurrent control,* (2) *dose-comparison concurrent control,* (3) *no treatment concurrent control,* (4) *active treatment concurrent control*, or (5) *historical* control, those who have submitted research plans to the FDA have found that it is usually very difficult to convince the FDA to accept anything other than a randomized placebo-controlled study.

Later, in his career, Lasagna moved to the University of Rochester and eventually Tufts University and founded the Center for the Study of Drug Development,[4] where multiday courses were taught to pharmaceutical industry researchers on designing clinical trials. RCT designs with restrictive inclusion and exclusion criteria were recommended to help assure a drug or device that would get a clean safety profile in the labeling. These restrictive criteria also minimized variation in the population being studied, thus minimizing sample size and maximized speed getting drugs on the market. Internal validity of these RCT studies were high, but once the drug was on the market, the labeling often did not help prescribers understand how the drug would actually work in patient groups that were not studied (elderly, young, minorities, those with comorbidities, childbearing age females, etc.). The regulations were followed, but the 1962 law's intent of "effective in use" might not have been achieved.

In some cases, the FDA realizes the use of a RCT is just not practical, and the clearest example given is the satirical article by Smith and Pell[10] reviewing the need for the RCT of parachutes for those who are gravitationally challenged. The RCT was not conducted to test the first-ever use of the rabies vaccine when a boy was bitten by a rabid dog, or with other therapies that clearly have an effect, where placebo would not be ethical, and where statistical methods with observational data are preferred.[11]

Other examples of where a traditional RCT is not feasible is for studying extremely rare events, where a sample size would need to be very large, and the cost or the time required to complete the trial may not be justifiable. For example, to compare low-molecular-weight heparin versus aspirin for pulmonary embolism (PE) mortality prophylaxis in patients with high-risk lower extremity fractures, a total of 19,579 patients would need to be enrolled to detect a 20% difference in the PE mortality end point ($P < 0.05$; 80% power).[12] If the RCT budget only allowed for 13,000 patients per arm, the difference observed in the PE mortality end point would need to be 33% (1.83% mortality vs. 1.46% mortality). The cost of a 13,000 patient per arm study would be approximately $11 million and could take 6 years to complete.[13] A controlled real-world evidence (RWE) study adjusting for known PE mortality risk factors could be more feasible because it can study a larger number patients with varying levels of risk and assess if there are mortality differences by treatment and by risk level. The results can be obtained sooner and at a lower cost. The advantage of knowing the research results sooner is the potential morbidity and mortality, which may be avoided if the optimal treatment is given to patients by their risk profile.

The FDA maintains a preference for placebo control; however, if a partly effective treatment exists, they acknowledge the ethical barrier to enrolling patients into studies of an unproven product, especially moderate-to-severe patients who would not want to take a chance of being treated with placebo in a trial. This movement of placebo control being unethical (for conditions with partly effective treatment exist) was

initially more common in Western Europe. RCTs moved to other second-world countries where the same ethical issues did not prevent enrollment, potentially further threatening the external validity of the study results to Western populations where lifestyles that can influence the condition are different. This also changed the types of patients who would enroll in some RCTs, further decreasing external validity. These and other ethical and scientific implications of the globalization of clinical research have been raised.[14]

The tight monitoring, encouragement, and coaching of clinical trial participants by "study nurses" further may decrease the external validity of the findings due to the Hawthorne effect. Organizational psychologists from Harvard were conducting productivity studies in the 1930s at the Hawthorne Plant owned by Western Electric in Cicero, IL. They noticed that because the employees were being isolated in an experimental suite and intensely measured, their productivity improved almost regardless of the intervention. As a result, they discontinued productivity research conducted in a tightly monitored subgroup of the workforce housed in a specialized environment. Further studies of workforce productivity by this team were conducted in the real-world factory setting.[15] In a recent evaluation of how the Hawthorne effect is being considered in medical research today, the authors reported that "researchers' lack of knowledge of this phenomenon is evident, despite evidence that the Hawthorne effect may cause over-optimistic results or false-positive bias."[16]

The 1962 law that was born with the hope of lowering drug prices and to providing patients and prescribers with data on the effectiveness of a product or device, morphed into a set of FDA regulations with some unintended consequences. Drugs and devices were being approved for use as being safe and effective, yet they were being used off label in patient populations and subgroups where safety and efficacy had not been studied. Often, the off-label uses comprise the majority of the product use. Numerous authors have identified limitations of the RCT.[17] The high cost of conducting RCTs with high-degree internal validity is often cited as a reason for the continued increases in drug prices and serves as a barrier to entry-preventing true price and product competition in the drug market, which Kefauver was trying to achieve.

The need for safety data was not fully satisfied by registration RCTs. With the growth of payer databases, FDA and claims data researchers began scouring the databases to identify adverse events, which were not identified in the highly controlled and rigidly monitored RCTs.[18] For some drugs or devices, safety monitoring became required after marketing.

It can be difficult to enroll patients into RCTs. Payer databases became a useful real-world data (RWD) tool for identifying patients who would qualify for clinical trials, and these patients were then referred to clinical trial sites. Registries were also developed as a means to funnel subjects into RCTs. RWD sources were tools being used to feed the RCT enrollment requirements. This ability to find patients for clinical trials enabled trials with very restrictive inclusion and exclusion criteria to enroll faster.

New Laws Passed to Fill Gaps in Medical Knowledge

Although the growth of the RCT continued in the early 2000s, the need for evidence to help clinicians make comparative effectiveness decisions grew in a health care system demanding greater efficiency. As a result, under President Obama, the American Reinvestment and Recovery Act of 2009 (ARRA, P.L. 111-5) provided a total of $1.1 billion for comparative effectiveness research and required the Institute of Medicine (IOM, now the National Academy of Medicine) to report recommendations for national comparative effectiveness research priorities. The following year, the Patient Protection and Affordable Care Act of 2010 (ACA, P.L. 111-148) authorized the establishment of a private, nonprofit, tax-exempt corporation called the Patient-Centered Outcomes Research Institute (PCORI), which often uses RWE to conduct comparative effectiveness research. Under President Trump, The Further Consolidated Appropriations Act, 2020 (P.L. 116-94, Division N, §104), extended funding for PCORI through FY2029.[19]

As one example of how these new funding options provide new forms of medical evidence, PCORI is now funding development of hybrid study designs to address the validity and generalizability issues caused by randomization, Hawthorne effect, and placebo control in RCTs.[20] Hybrid designs combine information about treatments and outcomes from clinical trials with a RWD on patients who receive the treatment.[20] Patients who usually receive a treatment in real-world settings may be different from patients who take part in randomized trials. First, the patients who use a treatment may change over time. Second, the way doctors prescribe treatments often differs from the way researchers test treatments in trials. The PCORI study team is working to improve the design of hybrid studies by developing methods that account for these 2 issues.

After a drug or device is approved for a specific indication and patient group, prescribers can choose to use it for a different indication and in other patient groups. In this case, the drug lacks the "effective in use" data, which the Kefauver-Harris amendment in 1962 attempted to establish. The payer databases and the electronic medical record (EMR) databases that have been developed after 1962 can potentially be used to identify and quantify the level of unmet need in a population or the extent to which the need is being satisfied by an existing drug or device. Although the regulations relevant to the Kefauver Harris amendments were written and modified over the past 60 years, there still is no requirement that utilization rates or real-world efficacy data be provided for the nonapproved indications of a drug or a device.

This growing safety and efficacy data gap between the population in which a product is studied and the population in which a product is actually used, in part, gave rise to the passage of the 21st Century Cures Act, signed into law by President Obama on December 13, of 2016.[21]

SEC. 3022. REAL-WORLD EVIDENCE …

1. In General: The Secretary shall establish a program to evaluate the potential use of real-world evidence (1) to help to support the approval of a new indication for a drug

approved under section 505(c); and (2) to help to support or satisfy post approval study requirements.
2. REAL-WORLD EVIDENCE DEFINED: In this section, the term "real-world evidence" means data regarding the usage, or the potential benefits or risks, of a drug derived from sources other than RCTs.
3. PROGRAM FRAMEWORK: The framework shall include information describing—the sources of real-world evidence, including ongoing safety surveillance, observational studies, registries, claims, and patient-centered outcomes research activities; …

The guidance "Use of Real-World Evidence to Support Regulatory Decision-Making for Medical Devices: Guidance for Industry and Food and Drug Administration Staff," was issued on August 31, 2017 under President Trump.[22]

Section 3022 of the 21st Centuries Cures Act and the regulations that followed allow for label expansion (new approved indications of existing approved drugs or devices) using RWE. The obvious advantage of this information is the increase in generalizability or external validity of the data, that is, if the data comes from real-world use of a product or device, it is very likely to apply to the real-world use of that product or device, as the American Medical Association argued in the early 1960s. This approach is consistent with the "effective in use" wording of the 1962 law. This new avenue for submitting effectiveness data to the FDA (and to scientific journals) should start to fill the information gap that has grown as trials became increasing strategically designed to demonstrate safety and effectiveness in a narrow or homogenous subgroup of patients.

The patient groups who are most likely to benefit from such RWE studies are those who are least likely to qualify for inclusion into a tightly designed RCT. These are usually those patients with concomitant illnesses (diabetes, cardiovascular disease, depression, etc.), the elderly, women of child bearing age, and others who are commonly excluded from randomized RCTs in an effort to 1) reduce heterogeneity in the efficacy measure or 2) reduce adverse event counts. Such a study design using large data bases also benefits the patients who have diseases or conditions with a low incidence, where a clinical trial for that indication is too expensive to justify the increase in sales, which may come from that rare condition.

The development of propensity scoring techniques to create balance between cases and controls greatly increase the internal validity of such RWE study designs. Case–control designs were available when President Kennedy signed the law requiring efficacy; however, those who wrote and shaped the regulation over the past 60 years often favored RCTs over epidemiology designs.

The growth in use of EMRs incentivized by the subsidies to medical practices for EMR as part of the Affordable Care Act in 2010 and also facilitated the ability to conduct valid studies of effectiveness and safety without the use of the labor-intensive RCT. EMRs have the potential to assure internal validity and increase external validity because most medical practices can now contribute data to collect evidence of safety and effectiveness without significant additional effort.

The Tufts Center for the Study of Drug Development, published the results of their benchmark RWE survey in 2018.[23] They reported that the RWE function was expected to grow 25% from 2017 to 2020 in the pharma and biotech industries. They concluded that RWE "will fill the gaps that are critically needed for drug development and safety." RWE is now listed as 1 of the Center's research properties.[24]

CONCLUSION

The "gold standard" method for collecting evidence regarding the safety and efficacy of a product or a device is not as clear as it was thought to be 60 years ago when Dr Louis Lasagna and others were advising the FDA on how to write regulations to support the Food Drug and Cosmetic Act amendment signed by President Kennedy in 1962. New tools and technologies available to the medical community have become available, and the call for a more ecumenical approach to clinical research have grown.[25] Statistical techniques for identifying a control group have improved. Our ability and need to understand rare diseases and conditions has increased. Our appreciation of patient subgroup differences (gender, age, ethnicity/race, lifestyle) and pharmacogenomic differences is forcing us to look past the RCT result in a homogenous population and ask for the relative safety and effectiveness of products and devices for many potential indications in many specific patient types. The RCT has successfully brought us from an era before efficacy data were required. The new RWE techniques may be the future standard because society and the medical community demands safety and effectiveness data for those patients and conditions who were underserved by the RCT.

REFERENCES

1. Greene JA, Podolsky SH. Reform, regulation, and pharmaceuticals—the kefauver-harris amendments at 50. *N Engl J Med.* 2012;367:1481–1483.
2. 21 U.S code § 355- new drugs [www.law.corness.edu]. Available at: https://www.law.cornell.edu/uscode/text/21/355. Accessed October 21, 2020.
3. Haynes B. Can it work? Does it work? Is it worth it? The testing of healthcare interventions is evolving. *Br Med J.* 1999;319:652–653.
4. Morelli M. Lou lasagna, father of clinical pharmacology dies [www.tuftsjournal.edu], 2011. Available at: https://tuftsjournal.tufts.edu/archive/2003/september/features/visionary_remembered.shtml. Accessed October 21, 2020.
5. Krug N. L. Lasagna, doctor, 80 and expert on placebos [www.nytimes.com], 2003. Available at: https://www.nytimes.com/2003/08/11/us/l-lasagna-doctor-80-and-expert-on-placebos.html. Accessed February 23, 2020.
6. Medical Research Council. Streptomycin treatment of pulmonary tuberculosis. *Br Med J.* 1948;2:769–782.
7. Feinstein AR, Horowitz RI. Double standards, scientific methods, and epidemiologic research. *N Engl J Med.* 1948;307:1611–1617.
8. Bonchek LI. Sounding board. Are randomized trials appropriate for evaluating new operations? *N Engl J Med.* 1979;301:44–45.
9. U.S Code of federal regulations [www.govregs.com], 2020. Available at: https://www.govregs.com/regulations/expand/title21_chapterI_part314_subpartD_section314.126#title21_chapterI_part314_subpartD_section314.126. Accessed October 26, 2020.
10. Smith GSC, Pell JP. Parachute us to prevent death and major trauma related to gravitational challenge: systematic review of randomized clinical trials. *Br Med J.* 2003;327:20–27.

11. Diaz M, Neuhauser D. Pasteur and parachutes: when statistical process control is better than a randomized clinical trial. *Qual Saf Health Care.* 2005;14:140–143.
12. Sample size to detect a significant difference between 2 proportions. [www.epitools.com.au], 2020. Available at https://epitools.asuvet.com.au/twoproportions. Accessed October 26, 2020.
13. Comparing medicines that prevent blood clots for emergency room patients with serious fractures- the prevent clot study [www.pcori.com], 2020. Available at: https://www.pcori.org/research-results/2016/comparing-medicines-prevent-blood-clots-emergency-room-patients-serious. Accessed October 20, 2020.
14. Glickman SW, McHuthison JG, Peterson Ed, et al. Ethical and scientific implications of the globalization of clinical research. *N Engl J Med.* 2009;360:816–822.
15. Wickstrom G, Bendix T. The "Hawthorne effect"—what did the original Hawthorne studies actually show? *Scand J Work Environ Health.* 2000:363–376.
16. Abdulraheem S, Bondermark L. Hawthorne effect reporting in orthodontic randomized clinical trials: truth or myth? Blessing or curse? *Eur J Orthod.* 2018;40:475–479.
17. Brothwell LE, Greene JA, Podolsky SH, et al. Assessing the gold standard—lessons from the history of RCTs. *N Eng J Med.* 2016;347:2175–2181.
18. Ray WA, Griffin MR. Use of Medicaid data for pharmacoepidemiology. *Am J Epidemiol.* 1989;129:837–849.
19. Funding for ACA-established patient centered outcomes research trust fund (PCORTF) extended through FY2029 [www.everycsrreport.com], 2020. Available at: https://www.everycrsreport.com/reports/IN11010.html. Accessed October 20, 2020.
20. Developing methods and guidance to improve hybrid study designs [www.pcori.org], 2020. Available at: https://www.pcori.org/research-results/2018/developing-methods-and-guidance-improve-hybrid-study-designs. Accessed October 20, 2020.
21. Public law 114-255- Dec 13, 2016 [www.congress.gov], 2016. Available at: https://www.congress.gov/114/plaws/publ255/PLAW-114publ255.pdf. Accessed October 20, 2020.
22. Use of real-world evidence to support regulatory decision making for medical devices [www.fda.gov], 2017. Available at: https://www.fda.gov/media/99447/download. Accessed October 20, 2020.
23. Lamberti MJ, Kubick W, Awatin J, et al. The use of real-world evidence and data in clinical research and postapproval safety studies. *Ther Innov Regul Sci.* 2018;52:778–783.
24. Tufts center for the study of drug development [www.csdd.tufts.edu]. Available at: https://csdd.tufts.edu/research-agenda. Accessed October 20, 2020.
25. Jones DS, Podolsky SH. The history and fate of the gold standard. *Lancet.* 2015;385:1502–1503.