Comparative Effectiveness Research

Vinod E. Nambudiri[1,2] and Abrar Qureshi[2]

Journal of Investigative Dermatology (2013) **133**, e5. doi:10.1038/jid.2012.497

WHAT IS COMPARATIVE EFFECTIVENESS RESEARCH.... AND WHY DO IT?

Comparative effectiveness research (CER) aids clinicians faced with medical decision making by identifying the best strategies among a variety of available preventive, diagnostic, and treatment options. Differing from early-phase clinical trials—in which an intervention is compared with a placebo and assessed for efficacy—the goal of CER is to discriminate among clinical interventions on the basis of clinical effectiveness, cost-effectiveness, adverse effects, or other distinguishing factors.

As part of the American Recovery and Reinvestment Act of 2009, the US government allocated $1.1 billion for the funding of CER with two primary aims: "(1) to conduct, support, or synthesize research that compares the clinical outcomes, effectiveness, and appropriateness of items, services, and procedures that are used to prevent, diagnose, or treat diseases, disorders, and other health conditions; and (2) encourage the development and use of clinical registries, clinical data networks, and other forms of electronic health data that can be used to generate or obtain outcomes data" (Department of Health and Human Services, http://www.hhs.gov/recovery/programs/cer/index.html, accessed 15 September 2012). One motivation behind the funding of CER is stimulating the delivery of higher-quality health care in a more cost-effective manner. Through well-designed and executed studies, CER has the potential to greatly enhance the practice of evidence-based dermatology (Williams, 2011). Common methodological approaches to conducting CER include randomized controlled trials and systematic reviews. This article will review recent examples of CER study designs in the dermatology literature as well as statistical analyses used to interpret such designs.

METHODS OF CER

CER may be conducted through a variety of study methods. One approach is to perform a systematic review of existing literature addressing one clinical question. Systematic reviews are detailed analyses and evaluations of all the published data on a specific topic to date. The aim is to draw conclusions from the large volume of data that are assessed across multiple published studies to answer the question at hand. These reviews offer the opportunity to conduct statistical analyses of aggregated data—a so-called meta-analysis—to gain broader insights that any one study would not have been large enough to assess. The use of patient registries built around specific clinical conditions facilitates such research by aggregating data for further study and analysis.

Another approach to CER is to design a randomized controlled trial to answer a specific clinical question. Studies that randomize patients to receive one commonly used medication versus another constitute a fundamental exercise of CER. Under this method, participants are randomly assigned to two or more groups that differ only on the basis of exposure to the study variable addressing the clinical question (namely, the medications, procedures, or diagnostic tools being compared). The groups are followed for predetermined outcomes of interest to address the question at hand, and the results of

WHAT COMPARATIVE EFFECTIVENESS RESEARCH DOES

- Aims to discriminate among clinical interventions on the basis of clinical effectiveness, cost-effectiveness, adverse effects, or other distinguishing factors.
- Answers questions from the patient and provider perspective of "which therapy is better?"
- Provide insights for future health-care policy and clinical decision making.

LIMITATIONS

- Conducting randomized trials to provide the best evidence is often expensive, labor intensive, and time-consuming.
- Rare conditions or disease states may not have sufficient individuals available for enrolling in such studies.
- Interpretation of studies is contingent upon appropriate study design and methodology.

[1]*Department of Internal Medicine, Brigham and Women's Hospital, Boston, Massachusetts, USA and* [2]*Department of Dermatology, Brigham and Women's Hospital, Boston, Massachusetts, USA*

Correspondence: Vinod E. Nambudiri, 45 Francis Street, Boston, Massachusetts 02115, USA. E-mail: vnambudiri@bics.bwh.harvard.edu

the two groups are compared by statistical analyses. Patients may be randomized at the individual level, or whole groups of patients may be randomized to particular interventions in the "cluster randomized" approach. Although often considered the gold standard for clinical research, randomized controlled trials are expensive, labor intensive, and time-consuming and may be particularly difficult to conduct for studying rare diseases.

CER IN DERMATOLOGY: THERAPEUTICS

Several common dermatologic conditions may be initially managed with a variety of medication classes. In a patient presenting with moderate acne, topical macrolide antibiotics, topical retinoids, topical benzoyl peroxide, and systemic antibiotics may all be considered part of the initial therapeutic regimen; similarly, for a patient presenting with mild to moderate atopic dermatitis of the face, topical corticosteroids or topical calcineurin inhibitors may be considered. Within each of these broad classes of medications, several treatment choices exist. Large randomized trials comparing multiple treatments head to head for a single condition—such as acne (Ozolins *et al*., 2004) or head lice (Chosidow *et al*., 2010)—offer important lessons for therapeutic agent selection by demonstrating significant differences in clinical effectiveness across treatments. The important point is that study participants must be randomly assigned to two or more treatments in the same study, using the same study design, double blind, with efficacy measured using the same disease severity indices. Surgical therapeutics may also be compared for effectiveness via randomized controlled trials, as has been done to assess surgical excision versus Mohs micrographic surgery for basal-cell carcinoma of the face (Mosterd *et al*., 2008). "Real-world" studies in which participants are patients treated in private-practice as well as academic settings, using on-label medications to manage the same disease process, are also considered within the scope of CER. The study design is cross-sectional, in which patients with the same clinical condition are treated with a range of therapeutic interventions and assessed for clinical response in a nonrandomized manner, as has been done with a variety of psoriasis treatments (Gelfand *et al*., 2012).

A large comparative effectiveness study published in the *Journal of Investigative Dermatology* in 2009 assessed two treatment regimens of the same steroid, clobetasol propionate, for disease control and event-free survival in patients with bullous pemphigoid (Joly *et al*., 2009). A total of 312 patients with moderate or extensive bullous pemphigoid were randomized to treatment with either high-dose clobetasol (40 g/day) or low-dose clobetasol (10–30 g/day). An important methodological component of any such trial is the *a priori* estimation of sample size, which calculates the number of subjects needed to detect significant differences in effects between interventions. The 2009 study was designed to have 80% power to detect a 33% difference in event-free survival between the two groups, with a one-sided log-rank test and type I error of 5%. Simply put, the statistical power of the study is the probability of finding a significant difference that does exist between the two groups; increasing the power of the study while holding other parameters equal will increase the number of experimental samples needed to reach the same level of significance. A type I error occurs when a difference between the two groups is claimed, although one does not actually exist. The probability of a type I error is known as α. Decreasing α—and thus reducing the probability of making such an error—while holding other parameters equal will require a larger sample size.

The bullous pemphigoid study cited above used the log-rank test for analysis of event-free and disease-free survival between patients in the two treatment groups (Figure 1). This test is used to assess differences between populations in the probability of an event over time, such as death or disease recurrence, and is often used for comparisons of survival between experimental groups (Bland and Altman, 2004). Such data are routinely plotted in Kaplan–Meier curves, which display time on the *x*-axis and percentage of surviving or unaffected individuals on the *y*-axis. Joly and colleagues reported no significant difference in overall event-free survival (patients unaffected by life-threatening adverse events or death) between the two treatment groups (*P* value = 0.95, Figure 1a). Significantly fewer side effects were seen in the lower-dose group. However, there was a significantly higher rate of disease relapse in subjects given the lower dose of steroids (*P* value = 0.012, Figure 1b). The authors concluded that the lower-steroid regimen demonstrated comparable

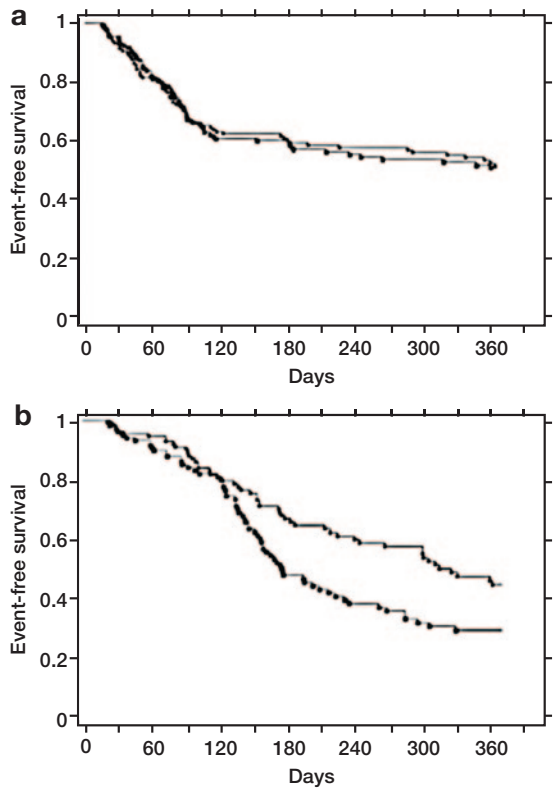

Figure 1. Kaplan–Meier curves. These curves demonstrate event-free and disease-free survival in patients treated with different topical steroid regimens for bullous pemphigoid. From Joly *et al*., 2009.