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

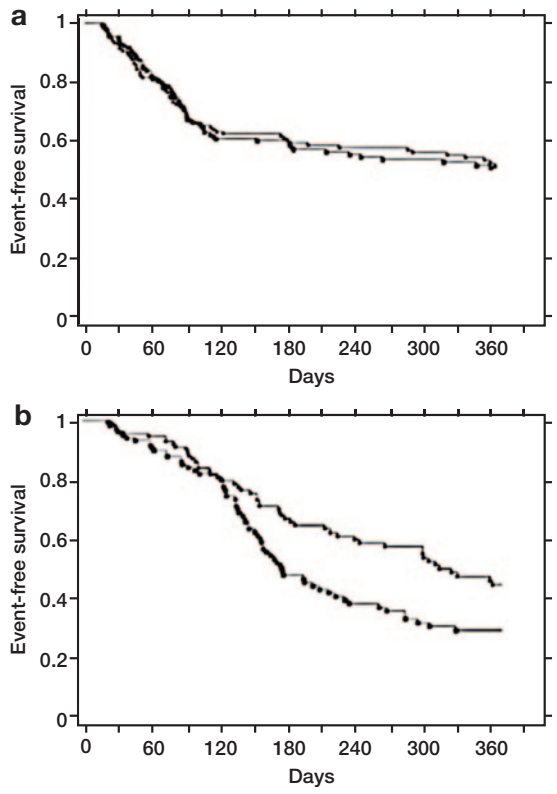

Figure 1. Kaplan–Meier curves. These curves demonstrate event-free and disease-free survival in patients treated with different topical steroid regimens for bullous pemphigoid. From Joly *et al*., 2009.

Table 1. Mean values of the Sk-29 and of the Sk-17 scores and their ICC in different dermatological conditions

		Symptoms			Psychosocial		
	n	Sk-29	Sk-17	ICC	Sk-29	Sk-17	ICC
Acne	193	33.2	37.3	0.939	32.1	33.6	0.927
Alopecia androgenetic	77	13.0	15.5	0.950	16.3	16.4	0.901
Alopecia areata	52	9.9	9.8	0.948	22.0	23.4	0.929
Bacterial infections	53	40.2	44.6	0.910	26.0	26.0	0.907
Balanitis	25	33.6	37.2	0.889	24.0	27.2	0.951
Benign skin neoplasias	175	13.9	15.7	0.930	10.0	8.2	0.907
Dermatitis	249	45.1	50.2	0.942	25.9	26.6	0.937
Hair loss	27	17.8	21.1	0.968	22.3	21.2	0.941
Lichen planus	32	34.2	38.4	0.942	20.3	16.8	0.914
Mycoses	116	28.1	31.8	0.954	17.4	17.0	0.925
Nail conditions	38	17.4	20.0	0.964	17.0	15.0	0.862
Nevi	306	10.3	12.3	0.925	8.7	6.7	0.911
Non-melanoma skin cancers	79	18.2	19.6	0.930	12.8	9.2	0.890
Pemphigus/bullous diseases	17	44.2	46.0	0.968	38.3	37.0	0.983
Pityriasis rosea	29	27.6	31.1	0.882	21.0	19.6	0.920
Pruritus	54	49.2	53.7	0.903	29.4	24.8	0.953
Psoriasis	220	47.9	54.2	0.946	32.5	33.0	0.941
Rosacea	60	33.3	33.2	0.951	26.2	26.1	0.953
Scabies and other ectoparasitic infections	34	51.5	57.0	0.942	40.7	40.8	0.935
Scars	19	29.1	33.2	0.971	33.2	33.3	0.942
Scleroderma/ connective tissue disorders	49	35.2	40.0	0.954	28.5	29.2	0.953
Seborrheic dermatitis	85	33.5	36.3	0.942	19.1	20.2	0.906
Urticaria	29	44.8	44.8	0.908	27.5	28.4	0.921
Viral infections	68	19.9	21.1	0.946	18.0	16.9	0.939
Vitiligo and other pigmentation disorders	54	12.0	11.3	0.916	19.5	20.3	0.921
Other dermatoses	60	39.1	41.3	0.948	29.0	30.3	0.920
Missing and other diagnoses	287	28.2	29.6	0.956	21.2	20.9	0.929
Overall	2,487	28.9	31.9	0.957	21.6	21.1	0.940

Abbreviations: ICC, intraclass correlation coefficient; Sk-17, Skindex-17; Sk-29, Skindex-29.

From Sampogna *et al.*, 2012.

clinical effectiveness with fewer side effects and lower overall costs that outweigh the slightly higher rate of relapse, a powerful lesson for both patients and their clinicians going forward. When addressing multiple treatment interventions in CER, selecting the appropriate statistical model and test is critical for a valid interpretation of the results and to limit research wastage caused by flawed designs or methodology (Williams and Delavalle, 2012).

CER IN DERMATOLOGY: DIAGNOSTIC TOOLS AND HEALTH-CARE TECHNOLOGIES

As a field in which clinical activities span physical diagnosis, medical and surgical treatments, long-term patient follow-up, and population health, clinical dermatology integrates several technologies into its routine practice. In addition to comparisons of pharmacologic treatments, CER can also be used to assess diagnostic instruments and other health-care technologies used in the practice of dermatology. There is a great need to develop outcomes in dermatology that are valid and reliably measure the diagnosis of a disease (e.g., screening questionnaires used in large epidemiology studies or diagnostic criteria used during a physical exam), disease severity (e.g., the Eczema Area and Severity Index), subphenotypes of disease (e.g., localized alopecia areata versus alopecia areata totalis), and health-related quality-of-life indices. Beyond development of indices and outcomes, CER is needed to better assess the use of measures in various settings (e.g., clinic vs. clinical trial). For example, efforts have begun to systematically review the process by which skin cancer is diagnosed by specialist and nonspecialist clinicians and to better understand the clinical impact of noninvasive technologies such as dermoscopy and photography on skin cancer diagnosis (Parsons *et al.*, 2011).

Another recent study compared the Skindex-29 with the Skindex-17, two health-related quality-of-life survey instruments that quantify the patient burden of dermatologic disease (Sampogna *et al.*, 2012). The Skindex-17 consists of a subset of questions that are derived from those in the longer 29-item questionnaire. Data from 2,487 patients who completed Skindex-29 surveys administered in a single institution were used to compute corresponding Skindex-17 scores. The mean values from these two sets of scores were then compared (Table 1). Sampogna and colleagues used intraclass correlation coefficients (ICCs) to compare the scores on the two questionnaires. The ICC is a statistical test used to calculate the reproducibility of measurements (or scores) of two different instruments measuring the same entity. The correlation coefficient is reported from 0 to 1. A high ICC represents a high degree of agreement between the scoring instruments. The two Skindex scores were found to have an ICC of 0.957 for questions regarding symptoms and an ICC of 0.94 for questions regarding psychosocial impact of disease. The high level of concordance observed between the instruments led the researchers to suggest that the shorter questionnaire may be effectively used to measure health-related quality of life, thereby reducing challenges associated with a longer survey, such as great respondent burden. Such research adds meaningful data by addressing the very metrics by which we assess dermatologic disease and also inspires further investigation of the effectiveness of our research tools and methodologies.

FUTURE DIRECTIONS FOR CER IN DERMATOLOGY

The current attention focused on CER by federal and international agencies looking to enhance the quality of medical-care delivery argues strongly for increased CER efforts within dermatology. Identifying meaningful future directions for such research should help translate into better,

evidence-driven, more effective patient care. Surveys of clinicians actively engaged in the treatment of patients with psoriasis have identified particular therapeutic interventions that these practitioners would like to see compared in future CER studies (Wan *et al*., 2012), providing guidance for additional work in the field. Recent studies comparing electronic-health and teledermatology visits for the management of chronic conditions such as acne (Watson *et al*., 2010) and atopic dermatitis (van Os-Medendorp *et al*., 2012) begin to address the value of technology-based care delivery in the practice of clinical dermatology. With proper planning and analysis, CER studies represent a powerful addition to the investigative dermatologist's toolkit for answering an array of complex questions. Current attention to better, more efficient, and lower-cost health-care delivery in the United States may be the burning platform for CER. With increasing demand for reducing variation and clinical process improvement, CER may finally receive the attention required to propel the next group of large studies. The ultimate challenge for the practicing clinician will be to translate these studies into better care for patients with dermatologic disease.

CONFLICT OF INTEREST

The authors state no conflict of interest.

SUPPLEMENTARY MATERIAL

Answers and a PowerPoint slide presentation appropriate for journal club or other teaching exercises are available at http://dx.doi.org/10.1038/jid.2012.497.

REFERENCES

Bland JM, Altman DG (2004) The logrank test. *Br Med J* 328:1073

Chosidow O, Giraudeau B, Cottrell MS *et al*. (2010) Oral ivermectin versus malathion lotion for difficult-to-treat head lice. *N Engl J Med* 362:896–905

Gelfand JM, Wan J, Callis Duffin K *et al*. (2012) Comparative effectiveness of commonly used systemic treatments or phototherapy for moderate to severe plaque psoriasis in the clinical practice setting. *Arch Dermatol* 148:487–94

Joly P, Roujeau JC, Benichou J *et al*. (2009) A comparison of two regimens of topical corticosteroids in the treatment of patients with bullous pemphigoid: a multicenter randomized study. *J Invest Dermatol* 129:1681–7

Mosterd K, Krekels GA, Nieman FH *et al*. (2008) Surgical excision versus Mohs' micrographic surgery for primary and recurrent basal-cell carcinoma of the face: a prospective randomised controlled trial with 5-years' follow-up. *Lancet Oncol* 12:1149–56

Ozolins M, Eady EA, Avery AJ *et al*. (2004) Comparison of five antimicrobial regimens for treatment of mild to moderate inflammatory facial acne vulgaris in the community: randomised controlled trial. *Lancet* 364:2188–95

Parsons SK, Chan JA, Yu WW *et al*. (2011) Noninvasive diagnostic techniques for the detection of skin cancers. In: AHRQ Comparative Effectiveness Reviews. Rockville, MD: Agency for Healthcare Research and Quality.

Sampogna F, Spagnoli A, Di Pietro C *et al*. (2012) Field performance of the Skindex-17 Quality of Life Questionnaire: a comparison with the Skindex-29 in a large sample of dermatological outpatients. *J Invest Dermatol* 133:104–109

van Os-Medendorp H, Koffijberg H, Eland-de Kok PC *et al*. (2012) E-health in caring for patients with atopic dermatitis: a randomized controlled cost-effectiveness study of Internet-guided monitoring and online self-management training. *Br J Dermatol* 166:1060–8

Wan J, Abuabara K, Troxel AB *et al*. (2012) Dermatologist preferences for treatments to compare in future randomized controlled comparative effectiveness trials for moderate to severe psoriasis. *Arch Dermatol* 148:539–41

Watson AJ, Bergman H, Williams CM *et al*. (2010) A randomized trial to evaluate the efficacy of online follow-up visits in the management of acne. *Arch Dermatol* 146:406–11

Williams HC (2011) Evidence-based dermatology: everyone's business. *Am J Clin Dermatol* 12:357–9

Williams HC, Delavalle RP (2012) The growth of clinical trials and systematic reviews in informing dermatological patient care. *J Invest Dermatol* 132:1008–17

QUESTIONS

1. Study designs used for comparative effectiveness research include which of the following?

A. Systematic review.

B. Randomized controlled trial.

C. Cross-sectional study.

D. All of the above.

2. Differences in survival between two treatment groups are best compared using which of the following statistical methods?

A. Paired *t* test.

B. Log-rank test.

C. ANOVA.

D. Fisher's exact test.

3. The ICC representing the best degree of agreement between two diagnostic tools among the following is:

A. <0.01.

B. 0.05.

C. 0.50.

D. 0.95.

Answers to the questions and an opportunity to comment on the article are available on our blog: http://blogs.nature.com/jid_jottings/.