Published in final edited form as:

J Comp Eff Res. 2012 ; 1(3): 263–270. doi:10.2217/CER.12.23.

Building useful evidence: changing the clinical research paradigm to account for comparative effectiveness research

Sheldon Greenfield* and Sherrie H Kaplan

University of California, Irvine, 100 Theory, Suite 110, Mail Code: 5800, Irvine, CA 92697, USA

Abstract

Comparative effectiveness research (CER) calls for substantial changes in the way clinical research is conducted, interpreted and practically applied in the USA, in order to produce useful clinical evidence. Departing from classic efficacy and effectiveness research, the evolving CER paradigm requires structural and substantive innovations that address three basic questions: what works? for whom? and in whose hands? Addressing these questions will require fundamental changes in the approach to clinical research that include: the use of active treatments (or comparators) versus placebos in the comparisons of treatments, innovative or 'alternative' research methods, the specification and *a priori* design of studies to account for important subgroups, accounting for the nested nature of healthcare delivery in design and analysis of CER, the simultaneous study of multiple treatments or treatment modalities, the study of multiple outcomes (benefits and harms) for each treatment compared, and the reassessment of the value of different study designs in the hierarchy of collective 'evidence'. In order to aid individual providers and patients in making informed, personalized treatment decisions, guided by the best evidence possible, CER studies must generalize to a broad range of subgroups reflecting the spectrum of patients, providers and health systems that populate real-world practice settings. Without expansion in the scope, conduct and subsequent interpretation of clinical research reflected in the issues outlined above, CER will fall short of its potential for informing evidence-based practice and personalized medicine. The current paradigm for conducting, interpreting and applying clinical research does not meet the needs of optimal generalizability and application to individual physician–patient efforts to identify the most effective treatment, and therefore does not support the basic requirements of CER. The proposed changes should neither require decades nor exorbitant budgets to achieve. Using two examples, prostate cancer, and comparisons of single medications, we illustrated how the proposed changes in clinical research, matching strategy to each application, might be addressed.

Keywords

comparative effectiveness research; enoxaparin; patient-centered outcomes research; prostate cancer; research methods

In order to produce usable clinical evidence, comparative effectiveness research (CER) is challenging the clinical research communities to substantially alter the approach to strategic planning of research, design and conduct of clinical research, and in addition, the

© 2012 Future Medicine Ltd

*Author for correspondence: Tel.: +1 949 824 5430 ext. 7286, Fax: +1 949 824 3388, sgreenfi@uci.edu.

Financial & competing interests disclosure

The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

interpretation and implementation of clinical research [1-4]. Departing from classic efficacy and effectiveness research, the evolving CER paradigm requires both structural and substantive innovations that address three basic questions: what works? for whom? and in whose hands? These questions are fundamentally consonant with those posed by the Patient-Centered Outcome Research Institute (PCORI) to characterize patient-centered outcomes research [101]. Answering them will require fundamental changes in the approach to clinical research.

These changes are highlighted by specific features that emerged from the Institute of Medicine (IOM) report [1]. They include use of active comparators versus placebos in the comparisons of treatments; innovative or 'alternative' research methods that allow for the simultaneous conduct of randomized controlled trials (RCTs) along with observational studies to evaluate both the benefits and harms of treatments in broadly defined populations at risk; the specification and *a priori* design of studies to account for subgroups of patients with differential potential for benefit from treatments; innovations in methods to account for the nested nature of healthcare delivery, not only for the impact of clustering by patient characteristics, but also by characteristics of providers (e.g., training, experience), and by features of healthcare systems themselves (e.g., decision support, information technology, provider assessment and feedback) in both study designs and analyses; the study of multiple treatments or treatment modalities simultaneously; the simultaneous study of multiple outcomes for treatments being compared that are of concern to patients and physicians in weighing treatment options; and reassessment of the value of different study designs in the hierarchy of collective 'evidence' that underpins optimal clinical practice.

Why is the current clinical research paradigm problematic? In addressing the question 'what works', treatments are infrequently studied in direct comparison to each other or to more than two treatments simultaneously because the current clinical research paradigm does not require it [5]. Similarly, multiple outcomes of interest to physicians and patients weighing treatment options (e.g., biomarkers, along with symptom relief, harms and quality of life measures) are infrequently studied with adequate power to compare treatments based on a spectrum of benefits and harms [6-8]. Physicians must therefore integrate the results of multiple different studies and make 'best guess' recommendations of specific treatment options to patients in their practices. Such decisions therefore frequently occur without support from sound empirical 'evidence'.

In addressing the question 'for whom', under the current paradigm, treatments found to be efficacious for the relatively homogeneous patient populations studied may not produce comparable benefits when used in the substantially more heterogeneous populations of both patients and physicians in real-world clinical practice settings [9-15]. The 'evidence' for evidence-based medicine (EBM) is therefore limited to a narrow and frequently unrepresentative group of patients from which physicians must make extrapolations to individual patients with varying characteristics and life circumstances [9,10,16]. The 'averaged effects' across larger studies with more inclusive patient populations often does not help to tailor information to important subgroups [12,16-18]. Such subgroups are rarely an *a priori* study design feature and therefore are not included in sufficient numbers to have adequate power to assess treatment effectiveness with confidence [19]. Therefore, in attempting to find the most effective and at the same time 'personalized' treatment options, physicians and patients are handicapped by evidence that does not cover the benefits and harms of specific treatments in direct comparisons to each other for broad enough patient populations.

To address the question, 'in whose hands', under the current paradigm, physicians and/or health systems are assumed to be a random effect, managed by 'within treatment'

randomization procedures. Some multicenter trials have been conducted to minimize the effect of differential administration of the intervention and/or the importance of higher or lower quality of care context on treatment outcomes [20]. With some designs (e.g., pragmatic trials, cohort studies), and with devices, some treatments and treatment modalities (e.g., robotic surgery for prostate disease), the quality of the provider may have a substantial impact on treatment effectiveness [21].

What are the potential solutions? First, solutions vary by the type of intervention being studied. Matching strategic research solutions to treatments under consideration should therefore be a primary principle of the evolving CER paradigm [4]. For example, strategic research planning should include recommendations for simultaneous conduct of studies representing the research design spectrum.

Classic efficacy approaches could be modified to include *a priori* modifications to study designs to address important subgroups in conjunction with observational studies to inform subgroup definitions and specifications [12,17,18,20,22]; the current requirement for classic RCTs could be modified to include the conduct of well-designed pragmatic trials and/or longitudinal observations studies, in order to maximize the number and types of treatments compared. Multiple ‘parallel’ well-designed, nonequivalent control group trials, and/or longitudinal observational studies with analytic attention to the importance of balancing groups by patient and potentially physician or practice site characteristics [23-25] could provide the required evidence needed by doctors and patients.

In addition, the clinical research and EBM communities must reconsider what constitutes valuable ‘evidence’ to inform routine treatment decision-making for specific treatment types [102,103]. If the evidence from RCTs is only applicable to small, homogeneous patient subgroups, physicians may be unable to practice EBM for the majority of patients in their practices. As the evaluation of physician performance broadens to include EBM and adherence to clinical practice guidelines based on EBM, absent pragmatic trials and/or observations studies, that assessment may be confined to a small sample of a physician’s practice and therefore misrepresent performance.

In summary, shifting of the clinical research paradigm must address: what works? (multiple treatment comparisons with multiple treatment outcomes); for whom? (design of studies to accommodate subgroups with differential potential for benefit from treatment); and in whose hands? (do different providers or clinical settings produce the same expected benefits from treatments).

The following cases have been used to illustrate the challenges and potential solutions related to the key features of CER as they apply to the study of different clinical problems, specifically the study of prostate cancer, and the comparison of medications.

Building new evidence: application of the key features of CER

The following examples were chosen to represent the varying solutions addressing the features of CER noted above. While not reflecting a comprehensive methodologic review of all possible solutions for each of the featured areas for both clinical examples, we proposed some practical alternatives to the current clinical research paradigm that, if implemented, could improve the evidence base and provide answers for physicians attempting to match treatments to patient needs.

CER in early prostate cancer

Background

Despite the advent of multiple different types of treatments for early prostate cancer, there is little compelling evidence recommending one treatment or treatment type over another. A very large RCT (ProtecT) is currently underway in the UK to evaluate the relative efficacy of radical prostatectomy, radical radiotherapy and active monitoring of prostate-specific antigen levels for men between the ages of 50 and 69 years [101]. This trial is investigating short and intermediate outcomes including disease progression, treatment complications, lower urinary tract symptoms, psychosocial impact of treatment, sexual function and quality of life, as well as all-cause and prostate-specific mortality. However, even such a large and well-designed study will not cover the spectrum of patients, providers and clinical circumstances that frame the multiple features of CER outlined above, and therefore will not be adequate to guide treatment decisions for patients, providers and treatments not represented in this study.

Problems

A RCT of nine treatments (or more as innovations in treatment become available) that includes multiple clinical sites and represents a broad spectrum of patients is neither practical nor likely to be undertaken in the USA for reasons not limited to funding restrictions (e.g., device marketing, patient preferences, physician preferences and industry-proffered financial incentives). Conducting useful CER for this clinical problem, especially using observational studies, in a relatively short time frame, could provide substantial benefit to patients and physicians struggling to make appropriate treatment decisions for this prevalent problem, but require innovations in study design and modifications of the valuing of such designs in the evidence hierarchy [102].

Solutions

First, what works? In the case of prostate cancer, active treatments versus placebos is not an issue of concern. For treatment of early prostate cancer, so many alternatives, many new and rapidly evolving, are available that the most useful data will derive from studies that simultaneously compare multiple treatments in real-world practice settings. What designs are best suited to multiple treatment comparisons?

Although until now not valued as highly in the evidence 'hierarchy', in this case, well-designed longitudinal observational (cohort) studies may best meet the requirements of CER and serve the needs of practicing physicians and their patients. With matching of patients to balance designs, either by *a priori* risk stratification [12,26,27] or by *post hoc* balancing of designs using propensity scores or instrumental variables [25], such studies could compare, for example, robotic-assisted prostatectomy to brachytherapy or prostate-specific antigen monitoring. Without randomization, risk stratification and/or design balancing efforts will be key to the interpretation of study results.

For which outcomes are these and other treatments for early prostate cancer more or less effective? Both long- and short-term outcomes, along with assessment of relative harms, contribute to the choice of effective treatment for early prostate cancer. Studies that compare treatments must also therefore consider mortality along with quality of life, sexual function and incontinence to be most useful for physicians and patients making treatment decisions [6].

Second, for whom? With observational study designs, the choice of treatment is in the hands of physicians and patients, not investigators. Bias by indication is therefore a key concern in

interpretation of study findings. Probably the most challenging issue in identifying effective treatments for early prostate cancer is matching patient subgroups using already well-characterized variables [12,19,26,27] with treatments that maximize short- and long-term benefits, balanced against their respective harms. For example, older patients with pre-existing heart disease or other major comorbidities may be unlikely to survive a sufficient length of time to benefit from an invasive procedure with significant side effects or ‘harms’ [6]. With data that support differential mortality rates and potential harms for these patients, physicians could provide evidence-based treatment recommendations. In order to provide such data, studies must be *a priori* designed to identify the subgroups with differential potential for benefit for specific treatments, and follow patients for side effects, complications and other harms.

If the longitudinal observation studies recommended above are to be most useful for CER, these studies must be either risk stratified by important subgroups and sufficient samples must be drawn to address power, and/or propensity scores or instrumental variables must be used to balance treatment ‘samples’. Such studies must follow patients for a sufficient length of time to include both short-term benefits (symptom relief) and harms (incontinence and sexual dysfunction), as well as long-term outcomes (e.g., all-cause and prostate-specific mortality). Practical identification of such subgroups *a priori* will require that registries of prostate cancer patients have adequate information. Although many such registries provide demographic, pathology, other laboratory and service utilization information, in order to optimize subgroup specification, additional information such as prevalence and severity of comorbid conditions, burden of illness or bothersomeness of symptoms from prostate disease, baseline functional status and self-reported quality of life must be added to routinely collected registry information [102]. *Post hoc* balance of treatment groups using propensity scores or instrumental variables will require that investigators collect these important subgroup variables during the course of the study and analyze appropriately.

Third, in whose hands? Because some of the available treatments and treatment modalities for early prostate cancer may be linked with physician experience, and because CER studies in this case are unlikely to be randomized to account for variations in physician or practice site quality of care, accounting for the ‘physician effect’ on treatment outcomes is essential, especially in interpreting data from longitudinal observational studies. Preliminary quality criteria for evaluating physician performance have been established for urology and should be included as mediating effects on treatment outcomes.

Updates

A well-designed multisite, longitudinal cohort study (the CEASAR Study) [102] funded in 2010 by the Agency for Healthcare Research and Quality, is currently ongoing. Using patients from five SEER tumor registries as an inception cohort, this study will prospectively compare different treatments and treatment modalities for major patient subgroups with differing potential for benefit from these treatments. Outcomes include cancer control, sexual function, other complications (e.g., urinary incontinence, bowel function), adherence to treatment and disease burden. Both risk stratification and use of propensity scores in analyzing treatment data are being utilized as part of this study.

Recommendations

To provide strong, if not definitive, evidence useful for discussions between doctors and their patients over a 5-year horizon, an initiative in early prostate cancer would include at least two kinds of studies, performed, hopefully simultaneously, with similar definitions of variables and similar definitions of study questions. One type would be homogenous small efficacy trials, and the other, as noted, large cohort observational studies. The latter would

directly and prospectively address two major elements: subgroups and quality of care. There may be multiple studies in each category. The efficacy trials are to show that the intervention works biologically under ideal conditions; the observational study is directed at harms (potency and continence) as well as benefits, for already characterized subgroups.

CER for individual medications such as enoxaparin

Background

The current landscape of studies of medication effectiveness is problematic for the conduct of CER for well-known reasons. For example, the current clinical research paradigm supports the use of placebos versus active comparators as the control group, uses RCTs as the sole design, uses ‘average effects’ in evaluating treatments that does not account for variations in effectiveness for important subgroups, for variations in harms for specific treatments and often uses single outcome measures that may be of limited use by patients and physicians in choosing among treatment alternatives.

Problems

In comparisons of medications, the appropriate comparator is often difficult to identify, particularly when multiple medications are already available for the same clinical problem. Under the current paradigm, the pharmaceutical industry does not have incentives to undertake such comparisons and would have difficulty operationalizing such studies. Even with appropriate comparators, the effect sizes for between-treatment differences will most likely be much smaller, requiring larger sample sizes or focusing on those subgroups most likely to benefit from treatments. To make optimal treatment decisions for the majority of patients in clinical practice settings, physicians need information on patients often excluded from studies under the current research paradigm, and for harms of treatments. Simply enlarging the size of trials without consideration for potential benefit from treatments for specific subgroups will not address this problem [28]. For newer or innovative designs such as pragmatic trials, control over intervention integrity also represents a problem for CER.

Solutions

First, what works? In order to provide greatest benefit for physicians and patients choosing among treatment alternatives, active comparators must be included in the new CER paradigm. Use of innovative designs such as well-designed pragmatic trials or longitudinal observational studies will help to overcome the generalizability limitations of evidence based primarily on RCTs that collectively exclude major patient subgroups (e.g., those with multiple comorbidity). These studies must also consider multiple outcomes as treatment benefits along with harms related to treatments to be of maximal use in clinical practice.

Second, for whom? Even with optimal active comparators and innovative study designs, patients with differential potential for benefit from treatments must be sampled or studied in adequate numbers to yield interpretable data in order to help guide treatment decisions in routine practice settings.

Third, in whose hands are medications most effective? In RCTs, intervention integrity, although much less problematic compared with surgical interventions, could still vary by practice site, physician training or experience. Although often considered a ‘random effect’, unless such studies are randomized within site, that assumption remains untested and may be a source of bias for CER studies. Further, patient adherence also compromises effectiveness of treatments. ‘Quality of care’, variations in treatment protocol due, for example, to patient adherence, and modifications made by physicians in response to patient preferences must be taken into account.

Updates

A study published in 2010, in the *Annals of Internal Medicine*, may provide an illustration of the level of flexibility and design innovation that could be useful for CER [13]. Hull *et al.* performed a large multisite RCT of enoxaparin compared with placebo in acutely ill medical patients. This trial originally did not intend to address two major problems in CER, namely a very effective comparator and differential effectiveness in key subgroups defined by well-researched patient characteristics. A review of the data after enrollment of 75% of the patients by the Data Safety and Monitoring Board (DSMB) revealed that the expected outcome events in the placebo control group were considerably lower than had been used in the power calculation. There were also some side effects, so the DSMB recommended discontinuation of the trial. When the interim data were analyzed further, it was discovered that the event rates in those patients with all three unknown ‘risk factors’ for venous thromboembolism (poor mobility, history of venous thromboembolism and history of cancer) had high event rates at the expected level. The study was ‘amended’ and continued with the 1310 high-risk patients instead of the 3685 originally sampled. The trial demonstrated effectiveness of the drug in this smaller high-risk sample. An accompanying editorial pointed out that a simple additive decision rule could have identified, *a priori*, the high-risk patients most likely to benefit before the study began [14]. The study could have been designed to focus on the patients who could benefit and reduce the chances of a null result, a not uncommon phenomenon in large randomized trials [10,16,29].

Kaplan *et al.* have developed multivariate scales, based on psychometric techniques, to identify patients more and less likely to respond to treatments for chronic diseases [10]. They built a composite variable comprised of a measure of comorbidity, a measure of functional status, a measure of depression, and a measure of the patient’s ability to work closely with the doctor to control the disease. These measures, comprising over 100 individual variables, all from patient report, formed an aggregate scale that met high standards for reliability. This scale, measuring ‘resilience’, divided into four groups, predicted HbA1c 1 year later in patients with diabetes, and also predicted a patient-reported measure of nonadherence [11,30,31]. The included variables were based on past research, which had shown them to individually predict patient outcomes. Based on the results, various drugs could be compared in a small number of preformed, adequately powered groups, defining differing levels of responsiveness to drug treatment, in randomized trials so that *post hoc* exploratory analyses based on single variables could be avoided. These results have been replicated in a recent study of four chronic diseases [32].

Thus, redesign of trials, both randomized and observational, especially those with active comparators, provide more individualized results than those resulting from using an ‘average’ effect, based on biology, disease severity, comorbidity and other salient patient characteristics. The interpretation of harms could be improved through such an approach because sicker patients are more likely to have adverse events independent of the treatment.

Recommendations

After the initial small RCTs, pragmatic trials and observational studies will be needed to determine the comparative effectiveness of a drug. Redesign of large classic drug trials where subgroup results are likely to differ from the average result because of biology, illness severity, or comorbidity and other patient factors is necessary. Some type of block design, with or without an adaptation of the study part way into the trial, could produce persuasive results without resorting to *post hoc* analyses, in a diverse patient sample, and in situations where usual care is better than expected. Two recent studies, based on the third category, comorbidity plus patient factors (‘resilience’), have been shown to predict patient adherence, 1 year a1c and utilization of services [12,32], forming four to five key groups that might be

expected to have heterogeneity treatment effects, independent of the treatment. A stratified design approach could be used for both pragmatic trials and for prospective observation studies where propensity scores could add to the balancing of groups within a block or overall. The interpretation of harms would also be helped by such approaches because sicker patients are more likely to have adverse events, independent of the treatment. The observational study, as noted, would also have the advantage of longer time to detect outcomes, larger samples to detect rare but important harms, multiple outcomes, and varying quality of care delivered in nontrial settings.

Acceptance: changing the evidence hierarchy in interpreting & synthesizing CER studies

There is little incentive to alter the current research paradigm to include pragmatic trials or longitudinal observational CER studies if those performing systematic reviews and developing clinical practice guidelines do not treat them as approximating RCTs or at minimum consider them to have sufficient internal validity to provide adequate evidence on which to base practice recommendations. The current criteria for assessing the quality of nonrandomized studies are inadequate because generalizability of evidence has not been a high priority to date. The report of the IOM on Systematic Reviews that has just been released continues to regard observational studies as useful primarily for the detection of harms, but not benefits [3].

The traditional evidence hierarchy should be modified to include well-designed and analyzed observational studies that include important predictors of outcomes and balance comparison groups using propensity scores and other techniques to compensate for nonrandomization. For example, a scenario that included two large RCTs, three carefully designed prospective cohort studies, all showing similar effect sizes and patterns for benefits and harms, addressing important subgroups and accounting for variations in quality of care (or fidelity to interventions) could lead to credible and useful clinical practice recommendations. Such would not be the case under the current research paradigm. Strategic research planning with flexibility in design and interpretation of studies could make the conduct of useful CER more efficient, shorten the time needed to produce usable results and lead to broader acceptance in clinical practice.

Summary

CER is challenging methodologists, statisticians and others who support the clinical and health services research community to reconsider and innovate in the areas of study design, data analysis, interpretation and hierarchy of evidence in order to help physicians and patients make tailored, effective treatment decisions. The evolving CER paradigm will need to make fundamental changes to address the basic questions of what works, for whom and in whose hands to be maximally useful and fulfill its potential.

Future perspective

With substantial increases in funding for CER and the commitment of PCORI to developing methods, tools and resources for conducting and evaluating it, the field is on the edge of significant transformation. Not only will CER continue to be defined and refined in contrast to the traditional research paradigm, but new techniques will (indeed, must) be developed especially for reliable use of carefully designed prospective observational studies to assess harms as well as benefits. Enhanced tools for subgroup analysis and the evaluation of heterogeneity of treatment effects will necessarily evolve for more personalized clinical decision-making. An extensive, high-quality data infrastructure will increasingly be set in

place to provide reliable data resources previously unavailable to researchers. Reassessment of the traditional research hierarchy will be both feasible and essential with the advent of these innovations. This transformation will underpin the generation of timely, valid and usable results for improved patient care.

Acknowledgments

The project described in this series of articles was funded by the National Pharmaceutical Council. S Greenfield has one NIDDK grant. The NIH award number is 5R18DK069846.

References

1. Institute of Medicine of the National Academies. Initial National Priorities for Comparative Effectiveness Research. Institute of Medicine of the National Academies; Washington DC, USA: 2009.
2. Institute of Medicine of the National Academies. Clinical Practice Guidelines We Can Trust. Institute of Medicine of the National Academies; Washington DC, USA: 2011.
3. Institute of Medicine of the National Academies. Finding What Works in Health Care. Institute of Medicine of the National Academies; Washington DC, USA: 2011.
4. Tunis SR, Benner J, McClellan M. Comparative effectiveness research: policy context, methods development and research infrastructure. *Statist Med*. 2010; 29(19):1963–1976.
5. Hochman M, McCormick D. Characteristics of published comparative effectiveness studies of medications. *JAMA*. 2010; 303:951–958. [PubMed: 20215609]
6. Chou R, Aronson N, Atkins D, et al. AHRQ series paper 4: assessing harms when comparing medical interventions: AHRQ and the effective health-care program. *J Clin Epidemiol*. 2010; 63:502–512. [PubMed: 18823754]
7. Varadhan R, Weiss CO, Segal JB, Wu AW, Scharfstein D, Boyd C. Evaluating health outcomes in the presence of competing risks: a review of statistical methods and clinical applications. *Med Care*. 2010; 48(6):S96–S105. [PubMed: 20473207]
8. Chou R, Fu R, Carson S, Saha S, Helfand M. Methodological shortcomings predicted lower harm estimates in one of two sets of studies of clinical interventions. *J Clin Epidemiol*. 2007; 60(1):18–28. [PubMed: 17161750]
9. Tinetti ME, Studenski SA. Comparative effectiveness research and patients with multiple chronic conditions. *N Engl J Med*. 2011; 364:2478–2481. [PubMed: 21696327]
10. Greenfield S, Kravitz R, Duan N, Kaplan SH. Heterogeneity of treatment effects: implications for guidelines, payment, and quality assessment. *Am J Med*. 2007; 120:S3–S9. [PubMed: 17403380]
11. Greenfield S, Billimek J, Pellegrini F, et al. Comorbidity affects the relationship between glycemic control and cardiovascular outcomes in diabetes: a cohort study. *Ann Intern Med*. 2009; 151(12):854–860. [PubMed: 20008761]
12. Kaplan SH, Billimek J, Sorkin D, Ngo-Metzger Q, Greenfield S. Who can respond to treatment? Identifying patient characteristics related to heterogeneity of treatment effects. *Med Care*. 2010; 48(6):S9–S16. [PubMed: 20473205]
13. Hull RD, Schellong SM, Tapson VF, et al. Extended-duration venous thromboembolism prophylaxis in acutely ill medical patients with recently reduced mobility: a randomized trial. *Ann Intern Med*. 2010; 153:8–18. [PubMed: 20621900]
14. Kent DM, Lindenauer PK. Aggregating and disaggregating patients in clinical trials and their subgroup analyses. *Ann Intern Med*. 2010; 153:51–52. [PubMed: 20621903]
15. Segal, JB.; Weiss, C.; Varadhan, R. Understanding Heterogeneity of Treatment Effects in Pragmatic Trials with an Example of a Large, Simple Trial of a Drug Treatment for Osteoporosis (White Paper). Center for Medical Technology Policy; MD, USA: 2011.
16. Kent DM, Hayward RA. Limitations of applying summary results of clinical trials to individual patients: the need for risk stratification. *JAMA*. 2007; 298:1209–1212. [PubMed: 17848656]

17. Kent DM, Rothwell PM, Ioannidis JPA, Altman DG, Hayward RA. Assessing and reporting heterogeneity in treatment effects in clinical trials: a proposal. *Trials*. 2010; 11:85. [PubMed: 20704705]
18. Rothwell PM, Warlow CP. European Carotid Surgery Trialists' Collaborative Group. Prediction of benefit from carotid endarterectomy in individual patients: a risk-modelling study. *Lancet*. 1999; 353:2105–2110. [PubMed: 10382694]
19. VanderWeele TJ, Knol MJ. Interpretation of subgroup analyses in randomized trials: heterogeneity versus secondary interventions. *Ann Intern Med*. 2011; 154:680–683. [PubMed: 21576536]
20. Hopkins LN, Roubin GS, Chakhtoura EY, et al. The carotid revascularization endarterectomy versus stenting trial: credentialing of interventionalists and final results of lead-in phase. *J Stroke Cerebrovasc Dis*. 2010; 19(2):153–162. [PubMed: 20189092]
21. Thorpe KE, Zwarenstein M, Oxman AD, et al. A pragmatic-explanatory continuum indicator summary (PRECIS): a tool to help trial designers. *CMAJ*. 2009; 180(10):E47–E57. [PubMed: 19372436]
22. Gage BF, van Walraven C, Pearce L, et al. Selecting patients with atrial fibrillation for anticoagulation: stroke risk stratification in patients taking aspirin. *Circulation*. 2004; 110:2287–2292. [PubMed: 15477396]
23. Glasziou P, Chalmers I, Rawlins M, McCulloch P. When are randomised trials unnecessary? Picking signal from noise. *BMJ*. 2007; 334(7589):349–351. [PubMed: 17303884]
24. Seeger JD, Kurth T, Walker AM. Use of propensity score technique to account for exposure-related covariates: an example and lesson. *Med Care*. 2007; 45:S143–S148. [PubMed: 17909373]
25. Stürmer T, Schneeweiss S, Rothman KJ, Avorn J, Glynn RJ. Performance of propensity score calibration – a simulation study. *Am J Epidemiol*. 2007; 165:1110–1118. [PubMed: 17395595]
26. Litwin M, Greenfield S, Elkin E, Lubeck D, Broering J, Kaplan S. Assessment of prognosis with the total illness burden index for prostate cancer (TIBI-CaP): validation of an instrument to aid clinicians in treatment choice. *Cancer*. 2007; 109(9):1777–1783. [PubMed: 17354226]
27. Daskivich T, Sadetsky N, Kaplan SH, Greenfield S, Litwin M. Less is more: severity of comorbidity and non-prostate cancer mortality in men with early-stage prostate cancer. *Arch Intern Med*. 2010; 170(15):1396–1397. [PubMed: 20696968]
28. Karanickolas PJ, Montori VM, Devereaux PJ, Schünemann H, Guyatt GH. A new 'mechanistic-practical' framework for designing and interpreting randomized trials. *J Clin Epidemiol*. 2009; 62:479–484. [PubMed: 18468856]
29. ACCORD Study Group. Effects of intensive glucose lowering in Type 2 diabetes. *N Engl J Med*. 2008; 24:2545–2559.
30. McHorney CA. The adherence estimator: a brief, proximal screener for patient propensity to adhere to prescription medications for chronic disease. *Curr Med Res Opin*. 2009; 25:215–238. [PubMed: 19210154]
31. McHorney CA, Victor Spain C, Alexander CM, Simmons J. Validity of the adherence estimator in the prediction of 9-month persistence with medications prescribed for chronic diseases: a prospective analysis of data from pharmacy claims. *Clin Ther*. 2009; 31(11):2584–2607. [PubMed: 20110004]
32. Cantrell R, Brenneman S, Kaplan S, et al. PDB10 assessing the ability of patient reported measures to discriminate proxies for treatment response. *Value Health*. 2011; 14(3):A92.

Websites

101. Patient-Centered Outcomes Research Institute. [14 March 2012] Questions to define patient-centered outcomes research. www.pcori.org/patient-centered-outcomes-research/
102. [28 February 2012] The ProtecT trial: evaluating the effectiveness of treatment for clinically localized prostate cancer. www.controlled-trials.com/ISRCTN20141297
103. [28 February 2012] Comparative Effectiveness Analysis of Surgery and Radiation (CEASAR) for localized prostate cancer. <http://clinicaltrials.gov/ct2/show/NCT01326286>

Executive summary

- The evolving paradigm for comparative effectiveness research requires both structural and substantive innovations that address three basic questions: what works? for whom? and in whose hands?
- Among the necessary innovations required are ‘alternative’ research methods for the simultaneous conduct of randomized controlled trials along with observational studies to evaluate both the benefits and harms of treatments in broadly defined populations at risk; the specification and *a priori* design of studies to account for subgroups of patients with differential potential for benefit from treatments; innovations in methods to account for characteristics of patients, providers and healthcare systems; the simultaneous study of multiple treatments or treatment modalities; the simultaneous study of multiple outcomes for treatments being compared that are of concern to patients and physicians in weighing treatment options. Specific examples of the application of some of these innovations are provided.
- The traditional evidence hierarchy should be modified to include well-designed and analyzed observational studies that include important predictors of outcomes and balance comparison groups using prematched subgroups, propensity scores and other techniques to compensate for nonrandomization.