

grace
principles

Good ReseArch for Comparative Effectiveness

The GRACE Initiative

www.graceprinciples.org

coordinator@graceprinciples.org



International Society of
Pharmacoepidemiology
endorses the GRACE Principles.



Good ReseArch for CComparative EEffectiveness OBSERVED

Prepared by the GRACE Initiative¹

Summary

The growth in therapeutic choices for many conditions has sparked interest in evaluating the comparative effectiveness of these alternatives. While randomized clinical trials (RCT) can provide high quality of evidence, non-interventional studies can provide data to fill the evidence gaps left by clinical trials, including providing information on subgroups of special interest, broader populations, certain conditions, treatment combinations and sequences, and understanding the effectiveness of actual use. However, methodological challenges for analysis and the interpretation of results, as well as the lack of accepted principles to assess quality have limited the practical use of observational research.

The GRACE principles describe a hierarchy of evidence for observational research on comparative effectiveness that can be used by decision-makers, as well as key elements of good practice including defining research questions and methods *a priori*; collecting valid, clinically relevant data; analyzing, interpreting and reporting data, including sensitivity analyses and alternative explanations for findings; and conducting these studies in accordance with accepted good practices. This living document is formed as three groups of questions that can be used to evaluate studies; it also provides guidance about what constitutes higher quality in terms of the accuracy of information generated and how to evaluate areas of uncertainty. No scoring system is provided or encouraged, since interpretation of these studies requires weighing all available evidence, tempered by judgment of their applicability to routine care.

Introduction

Comparative effectiveness (CE) has been defined as generating “the conduct and synthesis of research comparing the benefits and harms of different interventions and strategies to prevent, diagnose, treat and monitor health conditions in ‘real world’ settings.”¹ In this context, interventions can refer to the use of medical treatments,

¹ For a list of GRACE collaborators and supporters, see Attachment. Also see www.graceprinciples.org for more information, or write to coordinator@graceprinciples.org

including drugs, devices, and procedures. CE research is needed to support treatment decisions by patients and physicians as to what therapy to use, to guide formulary decisions, and to inform coverage policy decisions by payers.

It is widely accepted that the strongest form of evidence for the assessment of CE comes from randomized clinical trials (RCT).² Non-randomized studies have been relegated to lower tiers in commonly used hierarchies of evidence, largely because of their heterogeneity, the potential for bias in the results, and the challenges involved in their conduct and interpretation. However, these traditional evidence hierarchies were created primarily for the purpose of evaluating studies of the intended effect of a treatment. It has been argued that a different hierarchy of evidence is needed for evaluating studies that seek to explain how well therapies work in various subgroups.³ RCTs, as they are commonly implemented, are often limited in their ability to ascertain the effectiveness of healthcare products and services as they are actually used in the real world.⁴ For example, observed effectiveness depends in large part on the complex decision-making process of clinicians selecting therapies, decisions to stop or switch treatments based on tolerability, and so forth. Effectiveness also depends on the decisions and actions of patients, including whether they accept a course of therapy, how well a treatment is tolerated, the use of concomitant therapies, or treatment compliance once initiated. The impact of these issues on effectiveness may be difficult or impractical to assess in a traditional RCT, in which treatment assignments are protocol-driven and full compliance is assumed or actively monitored. RCTs may also be limited in situations where treatments or treatment practices are changing rapidly or where real-world use is broader than the types of patients that are typically accessible for and agreeable to a clinical trial. In addition, evaluations of long-term or sustained effectiveness of a specific therapy, which may change based on changes in tolerance and disease status, are generally not feasible under the typical time constraints of an RCT.

Awareness of these issues has sparked interest in more inclusive designs to evaluate the effectiveness of treatments for broader segments of the population.⁵ The value of observational research models in these settings is appreciated by many regulatory and advisory bodies involved in healthcare resource decision-making, but the methodological challenges must be addressed to facilitate meaningful analyses and reliable interpretations.

The methodological challenges in observational studies of CE primarily stem from the lack of randomization to treatment. This lack of randomization leads to concern about bias (systematic error) and confounding (a mixing of effects). A considerable amount of control over the potential for bias and confounding may be obtained in an observational CE study by design (exclusion, matching, or restricting the study groups to new drug users) and/or by analysis (restriction, stratification or mathematical modeling). However, sometimes the lack of clarity about the exact methodologies used, and the on-going debates as to best methods within the field lead to concern that the results are not valid.

Existing good practice guidance, like ISPE's Good Pharmacoepidemiology Practice, and ISPOR's good research practices for CE, provide broad, high-level guidance for conduct

and reporting, but do not focus on the evaluation of comparative effectiveness.^{6,7} The GRACE Principles set out to bridge that gap. These principles are offered in the form of a set of guided questions that may be useful as a high-level guide to evaluating good practices for observational studies of CE and are designed primarily for those who need to evaluate the rigor of such studies to inform decision-making regarding therapeutic alternatives, although they also may be useful for those who conduct these studies. These principles are consistent with good pharmacoepidemiologic practice,⁶ the Agency for Healthcare Research and Quality (AHRQ) handbook on Registries for Evaluating Patient Outcomes,⁸ ENCePP guidelines,⁹ and the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines for reporting observational studies.¹⁰ These principles also should be useful for those performing CE reviews, following either the Cochrane principles¹¹ or the AHRQ Guide for conducting comparative effectiveness reviews.¹² These principles have been introduced elsewhere; this document provides a more complete exposition on the introductory principles.^{13,14,15}

The GRACE principles, by themselves, are not sufficient to instruct the design, analysis and interpretation of comparative effectiveness research. Additional work needs to be conducted to guide the development of appropriate study designs for specific research questions, analytic strategies, and choice of data sets. The reader is encouraged to consult the literature and textbooks in epidemiology and outcomes research, and to participate in the International Society of Pharmacoepidemiology (www.pharmacoepi.org) and International Society of Pharmacoeconomics and Outcomes Research (www.ispor.org) to keep abreast of new developments in this field.

Good Research Practice for Observational Studies of Comparative Effectiveness

I. Was the study plan specified in advance of conducting the study?

A research plan should be developed before starting the study. The study plan should include clinically meaningful outcome measures that will assist patients and health professionals with treatment decisions or policy-makers with decisions about allocations of resources. This process describes key constructs that, in an ideal world, should be specified and evaluated. The study plan should be sufficiently detailed to allow replication of methodology. The following elements should be addressed:

1) The main purpose of the study, key research questions and target study population should be described.

Defining the research question/problem, including goals and objectives, at the outset of the study provides the context for structured data collection and a focused analysis plan. Objectives should describe the main patient population, outcome measures of interest, as well as the intended comparisons over a follow-up duration.

To the extent feasible, it is desirable to consider focusing CE studies on new users of the treatments of interest (also known as inception cohorts that only include people who have never been treated the medical intervention of interest). These cohorts are more likely to

provide a complete picture of the full benefits and risks of using new treatments, since this design avoids the “healthy user” effect that can result from studying people who are not new users.¹⁶ It is quite acceptable to include a broad range of patients to the extent affordable in order to enhance the information yield and generalizability of study results.

2) The diseases/conditions, comparators, treatment regimens, and the patient population of interest should be clearly defined, with consideration of the following:

- a. Diseases/conditions – diagnostic certainty, severity, time since diagnosis, significant co-morbidities, treatment history, etc.
- b. Comparators – comparison with one or many? Good CE research generally reflects the complexities of interventions as used in real practice settings. Comparisons to a number of real-world alternatives are generally preferable to a single comparator. Also consider to what extent the treatment and its therapeutic alternatives are already in use in the target population in sufficient numbers for meaningful analysis and interpretation
- c. Treatment regimens – for each comparator, consider the brand, dosage, method of delivery, duration of use, whether for a labeled indication, therapeutic alternatives currently in use, the likelihood that the necessary information will be accurately recorded and accessible, etc.
- d. Patient characteristics – age, sex, ethnicity, socioeconomic status, geography, access to medical attention (e.g., health insurance coverage), opportunity to be prescribed all of the medications that are being compared if their healthcare provider wanted to prescribe them, etc.

3) Measurement of effectiveness, safety and tolerability should be defined.

The outcome measure (endpoint) should be clearly stated, clinically meaningful, and appropriate for measuring effectiveness. The study plan should include clinically meaningful outcome measures that will assist health professionals and patients with treatment decisions, or policy-makers with decisions about allocations of resources. For example, differences in survival after invasive diagnostic procedures for acute myocardial infarction could be used to justify increasing the availability of cardiac catheterization labs,¹⁷ whereas decreases in a biomarker may not affect the risk of development of clinically apparent disease. Straightforward clinical outcomes are preferable to intermediate and composite endpoints; however, intermediate endpoints can be useful when there are good data that link the intermediate and long-term outcomes, and studying the long-term outcome is not feasible due to time or cost constraints.¹⁸ It is important that outcomes are meaningful and valid for all comparators, regardless of the intervention’s mechanism of action. This is especially important to consider when using surrogate markers of effectiveness (e.g., bone density as a surrogate for fracture).

Quantitative measurements of outcomes that are standardized, reproducible, and independently verifiable are preferable to clinical impressions and/or other measurements

that have not been validated or have substantial inter-observer variation, as long as those measurements are clinically important or are validated surrogate measures.

The method for estimating CE for the main objectives should be described, e.g., relative risk or risk difference. Although CE research usually entails a set of analyses that are adapted according to the patterns of data observed, the approach of specifying the main comparisons and analytic methods at the outset of a study will help assure skeptics that comparisons were not conducted iteratively until one was found to support a preconceived conclusion.

4) The intended study size should be described including a description of how that size was determined, what specific assumptions are being made, and how well these assumptions are supported?

II. Was the study conducted, analyzed, and reported in a manner consistent with good practice, and reported in enough detail for evaluation and replication?

All studies of CE should be conducted^{19, 20} and reported^{10, 21} in a fashion that makes the methods and practices clear. Consider the following:

1) How were the data collected or assembled, what checks were used to assure their validity, and were important data missing?

Primary data collection and secondary data used or collected for another purpose each have strengths and limitations that should be considered in the context of the study objectives. Both primary collection of data, i.e., data that are collected specifically for the purposes of the study, and secondary use of data, i.e., data that were collected for other purposes (such as administrative claims data and medical (health) records) require an understanding of how the data were collected, enrollment and coverage factors, pathways to care, quality assurance, other factors that may have affected the quality of the data and the validity of conclusions that may be drawn from their analysis.

Many of the methods that assure good clinical practice for RCTs²⁰ are appropriate for observational studies of CE. A main difference, however, is the trade-off between resources that are appropriate to devote to internal validity as compared with external validity. Whereas on-site monitoring to assure data quality is common for most, if not all, sites and patients in an RCT, observational CE studies generally have more sites and patients, and longer follow-up. Consequently, a larger proportion of budget is devoted to scope and duration in an effort to enhance broad generalizability which improves external validity, and less of the budget is devoted to individual data quality, which may diminish internal validity. Were budgetary constraints not a practical reality, this trade-off would not be required.

For prospective, primary data collection, studies should be designed to avoid affecting treatment patterns by creating an incentive for physicians to prescribe

certain treatments to fill study recruitment quotas, and study procedures should not be overly burdensome on patients or physicians. To the extent feasible, scales and measures for patient-reported outcomes should have been validated in populations similar to those under study and for similar methods of administration whenever possible. It is important to identify off-label use and treatments that may be used outside of their indication and to consider the feasibility of collecting data on these uses. For direct data collection, as an example, it can be difficult to recruit physicians prescribing a medication off-label without risking the appearance of promoting off-label use.

When using secondary data, a strong understanding of the data source and how the data were collected is essential and will help minimize errors in interpretation. For example, billing codes may not be reflective of the actual clinical condition, may be recorded inaccurately (e.g., coding errors), imprecisely (e.g., DRGs),^{22, 23} inconsistently, and/or under different constraints (e.g., one intervention might have been subject to pharmacy prescription limits, such as for migraine medicines).

Although all studies have missing data, some data related to exposures or outcomes of interest may be systematically missing (*not* at random). Some information may not be available in retrospectively collected data because the treatment or outcome of interest is not reimbursed by health insurance, is not accurately or completely recorded in sufficient detail for meaningful interpretation, or may be so sensitive that treatments are sought outside of the health system (e.g., treatment for drug addiction). Similarly, some drugs like IV antibiotics may not be recorded in prescription billing systems because they are dispensed in the doctor's office and are not distributed to patients through the pharmacy. Also, retrospective data rarely include information on over-the-counter products, although those products which may be important confounders. For both retrospectively and prospectively collected data, the accuracy and validity of the information on sensitive topics from the patient's perspective also must be considered (e.g., self-reported substance abuse or medication sharing) given patients may be reluctant to be truthful.

Whatever method of data collection is used, it is important to evaluate the patients actually selected for study (or enrolled) to determine how they compare to the target population in order to provide a clear context for evaluation of study findings.

- 2) Were the data analyzed by comparing patients who were similar in the characteristics that would have caused them to receive the treatment, and in their likelihood of benefiting from the treatment?

Patients should be analyzed in meaningful subgroups whenever patients who receive one treatment are believed to have a different baseline prognosis for the outcome of interest compared to patients who receive a comparator treatment to reduce the potential for confounding (e.g., patients with more severe/advanced disease generally have worse outcomes regardless of the therapy they receive). Such analytical techniques that should be used include conducting stratified analyses, creating

matched cohorts using propensity scores,²⁴ adjusting for prior events rates,²⁵ and using instrumental variables.²⁶

Intent-to-treat analyses can be used to assess prescribing practices, it is also important to examine actual use to the extent possible, including any means that can be used to quantify adherence and compliance.

3) Were alternative explanations for the findings considered and evaluated in terms of the potential impact on study interpretation?

Accurate interpretation depends on understanding the extent to which bias (stemming from factors that are related both to the decision to treat and to the outcome(s) of interest) may have distorted the results. Various types of bias to consider include the following:

- **Selection bias** refers to systematic differences among the groups being compared that arise from self-selection or physician-directed selection of treatments, or association of treatment assignments with other characteristics such as education, ethnicity, age, access to healthcare, etc.¹² Selective prescribing, or confounding by indication, describes the situation in which people with more severe forms of the disease/condition, or those who are resistant to other treatments, are more likely to receive newer treatments.
- **Misclassification** occurs when an exposure or outcome is incorrect or missing. Misclassification of drug exposure can result from the patient's incorrect recall of dose or poor adherence or treatment compliance. Studies using data sources that track prescription fills and refills are particularly vulnerable for treatments used on an as-needed basis (e.g., migraine medications) and for treatments dispensed in liquid or inhalable forms.²⁷
- **Detection bias**²⁸ applies to situations in which comparison groups are assessed at different points in time or using different methods or by assessors who may have knowledge of which treatment was used. Quantitative evaluations of outcomes that are standardized, reproducible, and independently verifiable are preferable to clinical impressions and/or other measurements that have not been validated or were not validated in the target study population.
- **Performance bias** refers to systematic differences in care other than the intervention under study. This bias often refers to differences that might affect adherence or persistence, or health practices such as diet, exercise, and smoking cessation. For example, a public health initiative promoting healthy lifestyles might be directed only at patients who have received one class of medical treatments, and the initiative, not the treatment, could be responsible for an observed benefit.

- **Attrition** refers to selective loss to follow-up. For example, if patients generally stop using a treatment with poor effectiveness but a small group of responders continue treatment, then the treatment could appear to have been more effective than it actually was. The effect of attrition can be addressed by characterizing those who drop out of studies, at what point, and why.²¹

Sensitivity analyses can provide a framework for evaluating the extent to which assumptions and common sources of bias may have explained any apparent differential effectiveness.

III. How valid is the interpretation of CE for the population of interest?

The main challenge for non-interventional studies of CE is to identify and disentangle systematic choices in prescribing that are related to the outcomes of interest. Various patient factors can lead to confounding (mixing of effects) of the estimated association between the treatment and outcome. A hierarchy of evidence can be applied to observational studies of CE to identify the situations which provide stronger evidence of quality, as well as other situations which may contribute useful information. The evidence hierarchy shown below is ranked according to the quality of evidence:

- Highest quality of evidence -- Determinants of treatment are not related to determinants of outcomes.

Treatment decisions are largely driven by reimbursements, such as different insurance plan formularies, rather than by patient characteristics or physician preferences. This situation enables an observational study to achieve an unbiased balance between comparisons groups because, in the previous example, choice of insurance (or residence, for national insurance plans) is generally unrelated to formulary decisions and treatment outcomes.

- Middle quality of evidence -- No consistent determinants of treatment, or determinants of treatments are largely known, or the risk of toxicity from treatment is unlikely to be related to the outcome(s) of interest.

Clinical equipoise occurs when a variety of treatments are frequently used and there is no good evidence for one treatment over another. Evidence for clinical equipoise might come from journal debates about how a patient should be treated, as an example. Strongly held, widely differing recommendations would signal that there is little evidence to support one treatment over another.

In other situations, a reliable understanding of the factors that drive physician treatment preferences exists and treatment determinants are independent of patient characteristics. In these situations, physicians always (or almost always) use the same product (e.g., anesthesia) and same course of treatment or same surgical approach. The experience of patients treated by physicians with differing strong, consistent practice patterns for these types of treatments could be compared,

although there still may be the opportunity for systematic error, e.g., high volume specialist generally have better outcomes than less experienced physicians.^{29, 30}

In some other situations, the risk of toxicity is known to be associated with a factor that drives prescribing patterns, but that risk factor appears to be unlikely to be related to the outcomes of interest. For example, warfarin is prescribed as an anticoagulant based on strict, widely used treatment guidelines to limit the risk of bleeding (e.g., ulcers). A study of stroke in patients on and off warfarin would provide meaningful evidence of CE if the only reason patients are not prescribed warfarin is because of a contraindication (ulcers) unrelated to the outcome of interest (stroke). The strength of this level of evidence depends on the likelihood that the determinants of treatment are truly unrelated to the outcomes of interest.

- Lower or indeterminate evidence quality -- Confounding and bias are likely to be present, but little relevant evidence is available

The lower quality level is reserved for those situations where the physician prescribing patterns for the treatment being compared have not been adequately described or a non-interventional CE study may reduce some uncertainty and provide some useful information, at an affordable cost, but it is difficult to understand to what extent unknown confounding factors could have artificially inflated the apparent relative benefit of one treatment compared to another. A particular challenge is the study of newly marketed drugs where there is substantial likelihood that new users will be either treatment resistant, or sicker than others, since stable patients do not usually change treatments. There are still some substantial advantages for observational studies, such as comparisons of surgical versus medical treatments, where RCTs tend to overestimate the real benefits achievable in routine clinical practice because of the technical expertise of surgeons in trials, and the rapid onset of treatment.¹⁷

There are other areas in which there are no randomized, head-to-head trials—pragmatic or otherwise – and any observational data derived from relatively rigorous studies with reasonable end-points and follow-up time should be considered. Consider autism for example. Like attention deficit disorder and other conditions in children, there is little research available about treatment effectiveness and great need for information.

Generally, unless an effect is observed that is much larger than would be expected or larger than could be explained by bias, it is unlikely that the study will contribute meaningfully to clinical decision-making. Although there is no unanimity about how large a relative benefit (e.g., relative risk of benefitting from a treatment) needs to be in order to be worthy of serious consideration as evidence for decision-making, some suggest that “as-treated” analyses showing a doubling in benefit or better should be given serious consideration.²⁵ However, some may set the bar as high as a five-fold benefit, which may be more useful for safety than for comparative effectiveness.²⁵

Conclusion

It is important to place information about CE into the public domain whenever possible, since it is important to add to the body of evidence that can be used to support decision-making to support inform therapeutic choices. The intent of the GRACE principles is to provide guidance for the execution and evaluation of observational studies of CE to enable decision-makers and other evaluators to be able to distinguish high-quality research.

It may be useful to report the results of observational studies of CE in the context of how well they support existing clinical trials data.^{31, 32} However, when the results of observational CE studies are not consistent with those of RCTs, especially when subgroups not studied in RCTs are included in the observational studies, it is not clear which interpretations are correct and which are not. Reporting observational CE studies, however they may be judged, may contribute to a better clinical and biological understanding of the disease, either by confirmation in a more targeted RCT or through advances in basic science.

References

1. U.S. Department of Health & Human Services. Federal Coordinating Council for Comparative Effectiveness Research: Report to the President and the Congress, June 30, 2009. (Accessed October 11, 2009, at <http://www.hhs.gov/recovery/programs/ce/ceannualrpt.pdf>.)
2. Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies and the hierarchy of research designs. *N Engl J Med*. 2000; 342: 1887–1892.
3. Vandembroucke J. Observational Research, Randomised Trials, and Two Views of Medical Science. *PLOS Medicine* 2008;5: e67:339-343.
4. Avorn J. In defense of pharmacoepidemiology – embracing the yin and yang of drug research. *N Engl J Med* 2007; 357: 2219-2221.
5. Rawlins, MD. De Testimonio. On the evidence for decisions about the use of therapeutic interventions. *Clinical Medicine* 2008; 8: 579-588.
6. Guidelines for good pharmacoepidemiology practice. *Pharmacoepidemiology and Drug Safety* 2008; 17: 200–208.
7. Berger ML, Mamdani M, Atkins D, Johnson ML. Good research practices for comparative effectiveness research: defining, reporting and interpreting nonrandomized studies of treatment effects using secondary data sources: The ISPOR good research practices for retrospective database analysis task force report – Part 1. *Value in Health* 2009; 12:1044-52.
8. Gliklich RE, Dreyer NA (eds). *Registries for Evaluating Patient Outcomes: A User’s Guide*. (Prepared by Outcome DEcIDE Center [Outcome Sciences, Inc. dba Outcome] under Contract No. HHS29020050035I TO1.) AHRQ Publication No. 07-EHC001-1. Rockville, MD: Agency for Healthcare Research and Quality. April 2007.
9. ENCePP. Checklist of Methodological Research Standards for ENCePP Studies. Draft for public consultation. 16 November 2009. Doc. Ref. EMEA/540136/2009.
10. Von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandembroucke JP. The Strengthening of Reporting of observational studies in epidemiology (STROBE) Statement: Guidelines for reporting observational studies. *Ann Intern Med* 2007; 147:573-577.
11. The Cochrane Collaboration. (Accessed November 27, 2007, at <http://www.cochrane.org/index.htm>.)
12. Agency for Healthcare Research and Quality. *Methods Reference Guide for Effectiveness and Comparative Effectiveness Reviews, Version 1.0* [Draft posted October 2007]. Rockville, MD. Available at: http://effectivehealthcare.ahrq.gov/repFiles/2007_10DraftMethodsGuide.pdf.
13. Dreyer NA, Berger M, Sullivan S, LeLorier J. Are Good Practices Principles for Observational Comparative Effectiveness Studies Needed? International Society for Pharmacoeconomics and Outcomes Research, Toronto, Canada, May 2008.
14. Dreyer NA, Schneeweiss S, McNeil B, Berger M, Walker A, Ollendorf D, Gliklich RE. *GRACE Principles: Recognizing high-quality observational studies of comparative effectiveness*. *American Journal of Managed Care*, In Press.

15. Dreyer NA, Rubino A, L'Italien GJ, Schneeweiss S. Developing good practice guidance for non-randomized studies of comparative effectiveness: a workshop on quality and transparency. *Pharmacoepidemiology and Drug Safety* 2009; 18: S123.
16. Ray W. Evaluating Medication Effects Outside of Clinical Trials: New-User Designs. *Am J Epidemiol* 2006; 158:915-920.
17. Stukel TA, Fisher ES, Wennberg DE, Alter DA, Gottlieb DJ, Vermeulen MJ. Analysis of Observational studies in the presence of treatment selection bias. *JAMA* 2007; 297:278-285.
18. Ramsey S, Willke R, Briggs A, Brown R, Buxton M, Chawla A, Cook J, Glick H, Liljas B, Petitte D, Reed S. Good research practices for cost-effectiveness analysis alongside clinical trials: the ISPOR RCT-CEA task force report. *Value in Health* 2005; 521-533.
19. Guidance for Industry: Good Pharmacovigilance Practices and Pharmacoepidemiologic Assessment. U.S. Food and Drug Administration. March 2005. Available online at: <http://www.fda.gov/cder/guidance/6359OCC.htm>.
20. Guidance for Industry: E6 Good Clinical Practice: Consolidated Guidance. April 1996. Available online at: <http://www.fda.gov/cder/guidance/959fnl.pdf>.
21. Tooth L, Ware R, Bain C, et al. Quality of reporting of observational longitudinal research. *Am J Epidemiol* 2005; 161:280-288.
22. Jollis JG, Ancukiewicz, DeLong ER, Pryor DB, Muhlbaier LH, Mark DB. Discordance of databases designed for claims payment versus clinical information systems. Implications for outcomes research. *Ann Intern Med* 1993;119: 844-850.
23. Lanza L, Walker AM, Bortnichak E, Dreyer NA. Peptic Ulcer and Gastrointestinal Hemorrhage Associated With Nonsteroidal Anti-inflammatory Drug Use in Patients Younger Than 65 Years A Large Health Maintenance Organization Cohort Study. *Ann Intern Med* 1995; 155:1371-13.
24. Glynn RJ, Schneeweiss S, Sturmer T. Indications for Propensity Scores and Review of Their Use in Pharmacoepidemiology. *Basic Clin Pharmacol Toxicol* 2006; 98: 253-9.
25. Tannen RL, Weiner MG, Xie D. Replicated studies of two randomized trials of angiotensin-converting enzyme inhibitors: further empiric validation of the 'prior event rate ratio' to adjust for unmeasured confounding by indication. *Pharmacoepidemiology and Drug Safety* 2008; 17: 671-685
26. Brookhart MA, Schneeweiss S, Rothman KJ, Glynn RJ, Avorn J, Sturmer T. Variable Selection for Propensity Score Models *Am J Epidemiol* 2006;163:1149-1156.
27. Strom, BL. Methodologic Challenges to Studying Patient Safety and Comparative Effectiveness. *Medical Care* 2007; 45: S13-S15.
28. Higgins JPT, Green S (editors). *The Cochrane Collaboration. The Cochrane handbook for systematic reviews of interventions, 2006.* (Accessed November 27, 2007, at <http://www.cochrane.org/resources/handbook/handbook.pdf>.)
29. Wennberg DE, Lucas FL, Birkmeyer JD, Bredenberg CE, Fisher ES. Variation in Carotid Endarterectomy Mortality in the Medicare Population: Trial Hospitals, Volume and Patient Characteristics. *JAMA* 1998; 279: 1278-1281.
30. Flum DR, Dellinger EP. Impact of Gastric Bypass Operation on Survival: A Population-based Analysis. *J Am Coll Surg* 2004; 199: 543-551.

31. Schneeweiss S, Patrick AR, Stürmer T, Brookhart A, Avorn J, Maclure M, Rothman KJ, and Glynn RJ. Increasing levels of restriction in pharmacoepidemiologic database studies of elderly and comparison with randomized trial results. *Medical Care* 2007; 45: S131-S142
32. GRADE Working Group. Grading quality of evidence and strength of recommendations. *BMJ* 2004; 328: 1490-1498.

Good Research for Comparative Effectiveness OBSERVED²

Leader

Nancy A. Dreyer, MPH, PhD, FISPE
Chief of Scientific Affairs
Outcome
Cambridge, MA

Coordinators

Michelle Leavy, MPH
Sr. Research Associate, Scientific Affairs
Outcome
Cambridge, MA

Pattra Wirojratana Mattox, MSc
Research Associate, Scientific Affairs
Outcome
Cambridge, MA

Collaborators

Marc L. Berger, MD
VP, Global Health Outcomes
Eli Lilly and Company
Indianapolis, IN

Jesse Berlin, ScD
VP, Pharmacoepidemiology
Johnson & Johnson Pharmaceutical R&D
Titusville, NJ

Michael N. Cook, PhD, MSc, MSPH
Director, Epidemiology
Global Safety Surveillance & Epidemiology
Pfizer
Collegeville, PA

Richard Gliklich, MD
President
Outcome
Cambridge, MA

Ken Hornbuckle, DVM, PhD
Sr. Epidemiology Advisor
Risk Management & Pharmacoepidemiology
Global Patient Safety
Eli Lilly and Company
Indianapolis, IN

Barbara J. McNeil, PhD, MD
Head, Dept. of Health Care Policy
Harvard Medical School
Boston, MA

Stephen Motsko, PharmD, PhD
Sr. Epidemiology Scientist
Risk Management & Pharmacoepidemiology
Global Patient Safety
Eli Lilly and Company
Indianapolis, IN

Dan Ollendorf, MPH
Chief Review Officer
Institute for Clinical & Economic Review
Boston, MA

John Parkinson, PhD
Director
General Practitioners Research Database
London, United Kingdom

Sebastian Schneeweiss, MD, ScD
Assoc. Professor of Medicine & Epidemiology
Harvard Medical School
Vice Chief, Pharmacoepidemiology &
Pharmacoeconomics, Dept. of Medicine
Brigham & Women's Hospital
Boston, MA

² The project received seed funding from the National Pharmaceutical Council.

Daniel E. Singer, MD
Professor of Medicine, Harvard Medical School
Director, Clinical Epidemiology
Unit & Research Fellowship Program
MGH General Medicine Division
Director, Epidemiology, Outcomes & Disease
Management Unit
MGH Clinical Research Program
Boston, MA

Til Stürmer, MD, MPH
Assoc. Professor & Director
Dept. of Epidemiology
University of North Carolina
School of Public Health
Chapel Hill, NC

Alexander M. Walker, MD, PhD
Principal
Whiscon
Wellesley, MA

Douglas J. Watson, PhD
Sr. Director, Medical / Scientific Liaison
Merck Research Laboratories
North Wales, PA

Supporters

José Luis Alloza, MD, PhD
Director, Alcalá Clinical Pharmacology
Project for Drug R&D
University of Alcalá
Madrid, Spain

G. Lawrence Atkins, PhD
Executive Director
Public Policy & Reimbursement
Shering-Plough Corporation
Washington, DC

Christina Farup, MD, MS
VP, Evidence Based Medicine
Novartis Pharmaceuticals Corporation
Hanover, NJ

Joy Frestedt, PhD, RAC, CCTI
President & CEO
Frestedt, Inc.
St. Louis Park, MN

R.M.C. Herings, PharmD, MPH, PhD, FISPE
Director, PHARMO Institute
Assoc. Professor of Pharmacoepidemiology
Dept. of Health Policy & Management
Erasmus MC
Rotterdam, The Netherlands

International Society of
Pharmacoepidemiology
Bethesda, MD

Mark V. Johnston, PhD
Professor, Occupational Therapy
University of Wisconsin-Milwaukee
Milwaukee, WI

Kristijan Kahler, PhD, RPh
Director, HE&OR, ABGU & NS
Evidence Based Medicine
Novartis Pharmaceuticals Corporation
East Hanover, NJ

Jeff Schein, DrPH, MPH
Sr. Director, Outcomes Research
Internal Medicine
Ortho-McNeil Janssen Scientific Affairs
Raritan, NJ

Robin Seitzman, MPH, PhD
Director of Epidemiology, Scientific Affairs
Outcome
Cambridge, MA

Hemal Shah, PharmD
Exec. Director, Health Economics &
Outcomes Research
Boehringer Ingelheim Pharmaceuticals, Inc.
Ridgefield, CT

J. Michael Sprafka, MPH, PhD
Vice President
Kendle Consulting
Cincinnati, OH