Research White Paper

Role of Single Group Studies in Agency for Healthcare Research and Quality Comparative Effectiveness Reviews

Agency for Healthcare Research and Quality
Advancing Excellence in Health Care • www.ahrq.gov

Research White Paper

Role of Single Group Studies in Agency for Healthcare Research and Quality Comparative Effectiveness Reviews

Prepared for:
Agency for Healthcare Research and Quality
U.S. Department of Health and Human Services
540 Gaither Road
Rockville, MD 20850
www.ahrq.gov

Contract No. 290-2007-10055-I

Prepared by:
Tufts Evidence-based Practice Center
Tufts Medical Center
Boston, MA

Investigators:
Stanley Ip, M.D.
Jessica K. Paulus, Sc.D.
Ethan M. Balk, M.D., M.P.H.
Issa J. Dahabreh, M.D., M.S.
Esther E. Avendano, B.A.
Joseph Lau, M.D.

AHRQ Publication No. 13-EHC007-EF
January 2013

This report is based on research conducted by the Tufts Evidence-based Practice Center (EPC) under contract to the Agency for Healthcare Research and Quality (AHRQ), Rockville, MD (Contract No. 290-2007-10055-I). The findings and conclusions in this document are those of the authors, who are responsible for its contents; the findings and conclusions do not necessarily represent the views of AHRQ. Therefore, no statement in this report should be construed as an official position of AHRQ or of the U.S. Department of Health and Human Services.

The information in this report is intended to help health care decisionmakers—patients and clinicians, health system leaders, and policymakers, among others—make well-informed decisions and thereby improve the quality of health care services. This report is not intended to be a substitute for the application of clinical judgment. Anyone who makes decisions concerning the provision of clinical care should consider this report in the same way as any medical reference and in conjunction with all other pertinent information, i.e., in the context of available resources and circumstances presented by individual patients.

This report may be used, in whole or in part, as the basis for development of clinical practice guidelines and other quality enhancement tools, or as a basis for reimbursement and coverage policies. AHRQ or U.S. Department of Health and Human Services endorsement of such derivative products may not be stated or implied.

Persons using assistive technology may not be able to fully access information in this report. For assistance contact EffectiveHealthCare@ahrq.hhs.gov.

None of the investigators have any affiliation or financial involvement that conflicts with the material presented in this report.

Suggested citation: Ip S, Paulus JK, Balk EM, Dahabreh IJ, Avendano EE, Lau J. Role of Single Group Studies in Agency for Healthcare Research and Quality Comparative Effectiveness Reviews. Research White Paper. (Prepared by Tufts Evidence-based Practice Center under Contract No. 290-2007-10055-I.) AHRQ Publication No. 13-EHC007-EF. Rockville, MD: Agency for Healthcare Research and Quality; January 2013. www.effectivehealthcare.ahrq.gov.

Preface

The Agency for Healthcare Research and Quality (AHRQ), through its Evidence-based Practice Centers (EPCs), sponsors the development of evidence reports and technology assessments to assist public- and private-sector organizations in their efforts to improve the quality of health care in the United States. The reports and assessments provide organizations with comprehensive, science-based information on common, costly medical conditions and new health care technologies and strategies. The EPCs systematically review the relevant scientific literature on topics assigned to them by AHRQ and conduct additional analyses when appropriate prior to developing their reports and assessments.

To improve the scientific rigor of these evidence reports, AHRQ supports empiric research by the EPCs to help understand or improve complex methodologic issues in systematic reviews. These methods research projects are intended to contribute to the research base in and be used to improve the science of systematic reviews. They are not intended to be guidance to the EPC program, although may be considered by EPCs along with other scientific research when determining EPC program methods guidance.

AHRQ expects that the EPC evidence reports and technology assessments will inform individual health plans, providers, and purchasers as well as the health care system as a whole by providing important information to help improve health care quality. The reports undergo peer review prior to their release as a final report.

We welcome comments on this Methods Research Project. They may be sent by mail to the Task Order Officer named below at: Agency for Healthcare Research and Quality, 540 Gaither Road, Rockville, MD 20850, or by email to epc@ahrq.hhs.gov.

Carolyn M. Clancy, M.D.
Director
Agency of Healthcare Research and Quality

Stephanie Chang, M.D., M.P.H.
Director
Evidence-based Practice Program
Center for Outcomes and Evidence
Agency for Healthcare Research and Quality

Jean Slutsky, P.A., M.S.P.H.
Director, Center for Outcomes and Evidence
Agency for Healthcare Research and Quality

Steven Fox, M.D., S.M., M.P.H.
Task Order Officer
Center for Outcomes and Evidence
Agency for Healthcare Research and Quality

Elise Berliner, Ph.D.
Task Order Officer
Center for Outcomes and Evidence
Agency for Healthcare Research and Quality

Acknowledgments

The authors gratefully acknowledge the following individuals for their contributions to this project:

Technical Expert Panel

Charles Anello, Sc.D., Sc.M.
Visiting Research Professor
Florida Atlantic University
Founder and Biostatistician
Applied Statistical Concepts, LLC
Wellington, FL

Sander Greenland, Dr.P.H., M.S., M.A.
Professor of Epidemiology
Professor of Statistics
University of California, Los Angeles
Los Angeles, CA

Timothy S. Carey, M.D., M.P.H.
Director, Cecil G. Sheps Center for Health Services Research
Professor of Medicine
University of North Carolina
Chapel Hill, NC

Robert L. Kane, M.D.
Co-Director, Minnesota Evidence-based Practice Center
Minnesota Chair in Long-Term Care and Aging
University of Minnesota
Minneapolis, MN

Role of Single Group Studies in Agency for Healthcare Research and Quality Comparative Effectiveness Reviews

Structured Abstract

Background. When systematically comparing procedures and therapies in the setting of a comparative effectiveness review (CER), the evidence base often includes single group studies, those that evaluate a single intervention given to all subjects included in the study design. The utility and limitations of single group studies to resolve clinical questions that are inherently comparative in nature should be described clearly.

Purpose. The purpose of this paper is to review the use and interpretation of single group studies in primary clinical research and to summarize current practices to using single group studies in CERs conducted by Agency for Healthcare Research and Quality (AHRQ) Evidence-based Practice Centers.

Methods. We extracted information about the use of single group studies in all published CERs prepared for the AHRQ Effective Health Care (EHC) Program through January 2012. Summary descriptive statistics across the different reports were tabulated.

Findings. Of 33 eligible reviews, 21 included single group studies. Ten used single group studies to report only harms, 2 reported only nonharm effects, while the remaining 9 reported both harms and nonharm effects. Ten of 21 did not provide the rationale for including single group studies. Of the 11 that did, the reasons cited included provision of supplementary long-term effect data on surgical interventions, hypothesis generation, and to address a paucity of existing comparative studies. Of the 12 CERs that did not include single group studies, 3 provided reasons for excluding these types of studies. They included specific concerns about confounding, avoidance of bias, and that sufficient data were expected from comparative studies. The terminology used to describe single group studies and their subtypes is not consistent across CERs.

Conclusions/recommendations. The reporting of inclusion or exclusion of single group studies in the EHC Program's CERs is suboptimal. Our review of published CERs to date indicates that single group studies are commonly included in CERs, but the rationale for including them is not consistently reported, and the methods relevant to their use are not clearly defined. Clarity and transparency in the rationale for including or excluding single group studies in CERs should be promoted. A working group should be convened to develop guidance on the circumstances under which single group studies should be included or excluded from a CER, to discuss how they should be integrated with evidence from other designs and how they may inform the strength of evidence assessment.

Contents

Tables

Figures

Appendixes

Background

Evidence from randomized controlled trials is often unavailable or insufficient to answer all questions posed in a comparative effectiveness review (CER). Thus, following a best-available-evidence approach,[1] systematic reviewers often use observational studies including a comparison group to examine the comparative effectiveness and safety of alternative therapeutic strategies. However, there are many instances where even observational studies with a comparison group are unavailable. Therefore, single group studies—those that evaluate a single intervention given to all subjects included in the study design—are often part of the evidence available to systematic reviewers conducting CERs.

We define a single group study as a study that consists of only a single group of subjects included in the study design, in which all subjects received a single intervention and the outcomes are assessed over time (i.e., not a cross-sectional study). These studies may be prospective or retrospective cohort studies. A number of study types would be included in this category, including investigations described as "single arm studies," case series, registries, "before-after designs," and time series studies. A classification scheme developed by Campbell and Stanley describes two single group studies consistent with our definition: the "one-shot case study" and the "one-group pretest–post-test design."[2] In the one-shot case study, a single group is studied only once after a treatment is applied. In the one-group pretest–post-test design, a pretest evaluation is followed by a treatment and then a post-test. For the rest of this paper, we will use the simplified term "single group study" when describing these designs in general.

Single group studies are often conducted in the setting of strong therapy preferences (e.g., hyperbaric oxygen therapy for arterial gas embolism[3]). This is especially true for transplantation studies of vital organs in the setting of rapid and fatal disease progression. For example, in patients with end-stage liver disease, the natural history of disease is so well known that it would be difficult to carry out a trial with an untransplanted study arm. Also, a field of clinical inquiry that is relatively new may not be sufficiently mature to rationalize a comparative hypothesis. For example, novel procedures or drugs are often initially evaluated in single group studies that are used to inform the design of a subsequent study with an internal comparison group.

Single group study designs are commonly used to monitor adverse events that may become evident only with long-term followup of large numbers of treated patients, which is not practical or efficient with other study designs. For example, phase 4 studies to monitor postmarketing adverse events and evaluations of therapies often include a single group of patients managed with the same treatment strategy and followed over time. Open-label extensions of clinical trials present another type of clinical investigation that often lacks an internal, concurrent comparison group. Although they are designed to follow patients for an extended period of time, they also usually evaluate a more highly selected population of patients who completed the randomized trial, tolerated the medication, and agreed to participate in the extension. Expanded access programs (or "compassionate use") allow the use of an investigational drug outside of a clinical trial to treat a patient with a serious or immediately life-threatening disease or condition lacking satisfactory alternative treatment options. These investigations commonly describe the experience of a single group of patients without a comparison group (for example, see Janne 2004[4]). Finally, registries of patients who have been exposed to a single drug or device may also be assembled for monitoring long-term sequelae without an internal comparison group. An example includes the coordinated effort to study newly introduced devices through the Interagency Registry for Mechanically Assisted Circulatory Support, established to capture

detailed clinical data on all patients receiving implantable ventricular assist pumps in the United States.[5]

Since single group studies do not include a direct, concurrent comparison group, their role in informing comparative effectiveness questions is not straightforward. Observational study designs in general suffer from a potential lack of exchangeability of exposed and unexposed subjects. In other words, the outcome in the untreated group may differ from what would have occurred in the treated in the absence of treatment (the "counterfactual outcome"). The absence of a direct, concurrent untreated comparator in single group studies presents an added challenge to identifying a proxy for the counterfactual, or an answer to the question: "What would have been the treated person's experience if there had been no treatment?" Extrapolations based on the expected outcomes in the "missing" untreated arm are required for inference about treatment effects. In fact, explicit and implicit comparisons are frequently made in single group studies even in the absence of a direct, concurrent comparator. The appropriate interpretation of these implicit and explicit comparisons and their potential utility in CERs must include consideration of the key assumptions underlying each single group design.

The ability of observational studies to answer questions about the benefits or intended effects of pharmacotherapeutic agents, devices, or procedural interventions has been a matter of debate.[6] Guidance has been developed for systematic reviewers for decisionmaking on the inclusion of observational studies in general in CERs.[6] However, to the best of our knowledge, the use of single group observational studies in CERs has not been specifically addressed in this methods guide or elsewhere. While the value of using single group studies to identify and quantify the occurrence of harms[a] of interventions is well recognized, the role of these studies in evaluating comparative effectiveness and safety is not well developed. Given that single group studies may comprise a substantial portion of the evidence base for a given clinical question, and in light of the challenges in their interpretation and relevance to questions that are comparative in nature, it is important to clarify whether they are useful in informing comparative effectiveness assessments, and if so, to clarify the assumptions required to support their use.

In order to illuminate the use of single group studies in CERs, we conducted an empirical review of current practices in using single group studies in CERs conducted by Evidence-based Practice Centers (EPCs) for the Agency for Healthcare Research and Quality (AHRQ). The summary findings should serve as an impetus for future work in reaching a consensus across EPCs as to when and how single group studies should be used in CERs specifically and systematic reviews in general. In addition to the empirical review, we also provided a narrative review section describing the common single group study designs and the key considerations and assumptions required for their interpretation to help guide comparative effectiveness reviewers who encounter this type of evidence.

[a] Includes adverse events of interventions as well as other harmful events that may be indirectly related to the intervention.

Methods

The purpose of the following empirical evaluation is to understand how EPCs have used single group studies in their CERs. Because there was no single term used consistently across EPCs to describe these types of studies, we reviewed the Methods section of each published CER to search for relevant descriptions (e.g., noncomparative cohort). These descriptions are provided in Appendix A. Crossover studies were excluded from this review because these designs are well-characterized and their properties have been extensively discussed in the literature.[7-9]

We identified all published CERs prepared for the Effective Health Care Program through searches of the Effective Health Care Web site (www.effectivehealthcare.ahrq.gov). The last search was conducted on January 10, 2012. We excluded CERs at the draft stage at the time of the last search because draft reports have not been peer reviewed and may not be representative of the final product. We included updates of a previously completed CER when the approach to single group studies differed from the original reports.

From each eligible report we only considered "comparative" Key Questions, those that explicitly compare benefits or harms between alternative management strategies. Two reviews pertaining only to diagnostic studies were excluded as they did not evaluate comparative treatment effectiveness.[10,11] These two reviews included mostly individuals that had received both diagnostic and reference tests.

Five team members piloted and modified the data extraction form. Thereafter, a single reviewer extracted information from each CER. We collected the following information for each report: title, year of publication, types of interventions compared, whether the authors planned to use single group studies, selection criteria for single group studies (minimum sample size, followup duration, or "other"), whether the results of different single group studies were synthesized separately or together with comparative studies, methods for evidence synthesis (qualitative or quantitative), methods for risk of bias (or quality) and strength of evidence assessment, results of the review regarding single group studies (total number identified, and ratio of single group to total number of studies included), and specific synthesis methods actually utilized in the report. When appropriate, information on the above items was extracted separately for treatment benefits and harms. Whenever possible we extracted the rationale for the methodological approach used. We tabulated summary descriptive statistics across the different reports. The Summary Table of Empirical Review Extractions is presented in Appendix A.

To gain insight into the methodologic issues concerning the use of single group studies in CERs, we convened a panel of experts to help guide this project. The panel members reviewed two drafts of this paper and also provided additional feedback in a teleconference.

Results

We identified a total of 40 CERs, of which 34 were original CERs and 6 were updates. The list of included CERs is in Appendix A. Because two CERs on diagnostic tests were excluded, we extracted information from the 32 original CERs evaluating interventions that were not exclusively focused on diagnostic tests, plus 1 update CER whose methods and inclusion criteria differed from those of its original CER. Of the 33 extracted CERs, 21 used single group studies in their review. Of the 21 CERs that used single group studies, 10 reviewed only pharmaceutical agents, 1 reviewed only surgical interventions, 1 reviewed radiotherapy treatments, and 9 reviewed multiple types of interventions. Ten (48%) used single group studies to report only harms, 2 (10%) used single group studies to report only benefits, while the remaining 9 (43%) used single group studies for reporting both harms and benefits (Figure 1). There was some indication of differences in the use of single group studies depending on the type of intervention being evaluated (Figure 2), with single group studies used least commonly in evaluating pharmaceutical interventions, and most frequently used for assessing only harms in evaluation of medical devices. However, this interpretation is limited by the small sample size.

Figure 1. The use of single group studies in 33 comparative effectiveness reviews

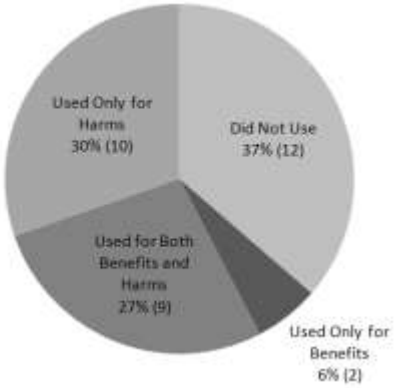
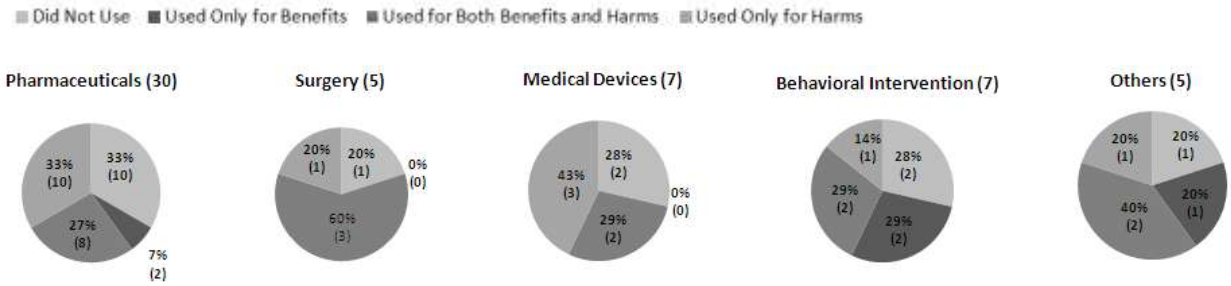


Figure 2. The use of single group studies in comparative effectiveness reviews by type of interventions assessed



It is noteworthy that not all of the 21 CERs that included single group studies provided an explicit rationale for doing so (Table 1). Of the 11 that did, the reasons cited included provision of supplementary long-term effect data on surgery, hypothesis generation, and addressing a paucity of existing comparative studies. Of the 12 CERs that did not include single group studies, three provided reasons for excluding these types of studies. They included specific concerns about confounding; avoidance of bias; and that sufficient data were expected from

comparative studies. We were not able to determine whether the decisions to include or exclude single group studies were made a priori or post hoc.

Table 1. Characteristics of reviewed comparative effectiveness reviews prepared by EPCs

Characteristics	Result n (%), Unless Otherwise Stated
Total number of CERs reviewed	33
Year of publication, range	2005–2011
Included single group studies	21 (64)
Stated rational for including single group studies	11/21 (52)
Stated rationale for including single group studies for benefits	6/11 (55)
Stated rationale for including single group studies for harms	8/19 (42)
Methods	
Planned to quantitatively synthesize single group studies	1 (5)
Planned to quantitatively combine single group studies and comparative studies	0 (0)
Specified the reasons for using single group studies in quantitative analysis	2 (10)
Planned to qualitatively combine single group studies and comparative studies	5 (24)
Describe their methods for assessing risk of bias/quality of individual studies	21 (100)
Explicitly describe their methods to assess the risk of bias/quality of included single group studies	12 (57)
Results	
Clearly provided the total number of studies included	20 (95)
Total studies included, median (25^{th}–75^{th} percentile)	128 (74–200)
Clearly provided the number of single group studies included	14 (67)
Total single group studies included, median (25^{th}–75^{th} percentile)	21 (8–71)
Assessed strength of evidence for the Key Question	18 (86)
Used single group studies to determine strength of evidence	8 (44)
Quantitatively synthesized single group studies for efficacy/effectiveness	0 (0)
Quantitatively combined single group studies with comparative studies for efficacy/ effectiveness	0 (0)
Quantitatively synthesized single group studies for harms	0 (0)
Quantitatively combined single group studies with comparative studies for harms	0 (0)
Described the syntheses of the effects from single group studies adjacent to the syntheses of comparative studies	5 (24)
Described the syntheses of the harms from single group studies adjacent to the syntheses of comparative studies	9 (43)

CER = comparative effectiveness review; EPC = Evidence-based Practice Center

The median number of single group studies included in the CERs was 21 (interquartile range 8, 71). Notably, none of the 21 CERs that included single group studies quantitatively synthesized single group studies for any outcomes, nor did they attempt to quantitatively combine results from these studies with those from comparative studies. Eighteen of the 21 CERs that included single group studies conducted an assessment of the strength of evidence, and single group studies were used to determine the strength of the evidence in 44 percent of these reviews.

Terms used to describe single group studies in the CERs included "cohort study," "observational study," "single arm study," "noncomparative observational study," and "case study" (Appendix A).

A summary table of the use of single arm studies in the CERs reviewed is presented in Appendix A, including the title of the reports and the types of interventions, whether single

group studies were used to address benefits or harms, the rationale for the inclusion of single group studies, and the terms used to describe single group study.

Discussion

Empirical Review

Our review of published CERs to date indicates that although single group studies are very commonly used in CERs, the rationale for including them is not frequently reported. In addition, the methods and approaches to how single group studies are used—either in isolation or integrated with comparative studies—are not clearly defined in the CERs. As an example, several CERs used single group studies to inform the strength of evidence, but the rationale or methodologies were not described. Furthermore, we observed that the terminology used to identify single group studies and their subtypes is not consistent across CERs. For example, single group studies were occasionally identified in CERs using terms that also describe observational studies with an internal concurrent comparison group (e.g., "cohort study," "observational study"). The lack of consensus on terminology for these designs and their subtypes makes the identification of single group studies in CERs challenging for readers.

It is notable that the majority (19/33, 58%) of the CERs used single group studies in the reporting of harms. Only about half of these CERs (9/19, 47%) described these harms data in the context of data from comparative studies. For those CERs that could not (or did not) juxtapose single group data alongside comparative data, it would be difficult to explore potential causal attributions of the reported harms.

Our empirical evaluation of the use of single group studies among EPCs in their CERs has several limitations. Only a single investigator extracted data from each CER, which could result in misclassification given the subjective nature of some of the elements abstracted. However, we pilot-tested the data extraction form as a group in order to operationalize the process of data abstraction and to promote agreement between readers. A second limitation of the data extraction process was the lack of consistent nomenclature used across EPCs to describe single group studies. This lack of consistent and clear terminology, as well as opacity in reporting on how these studies were used, made our identification of these designs and extraction of information on how they are used challenging. This limitation underscores our recommendation that development of consistent nomenclature for these study designs be prioritized, and that reporting on their use and application be made more transparent. Finally, our empirical review was conducted among the small sample of CERs that had been prepared by EPCs for AHRQ through January 2012 and may not be generalizable to the conduct of comparative evidence reviews in general.

Given ethical and practical constraints to many areas of clinical investigation, it is likely that single group studies will continue to be used to inform CERs, especially in the evaluation of novel technologies or in the examination of harms. Clarity and transparency in the rationale for including or excluding single group studies in CERs should be promoted. As a step forward, we have developed suggestions to improve the reporting of single group studies in CERs (Table 2). We further recommend that a working group across EPCs be convened to develop consistent terminology to identify single group studies, specify circumstances under which single group designs should or should not be included in a CER, discuss how data from single group designs should be considered together with other research designs, and how to incorporate data from single group designs into the deliberation of strength of evidence assessment.

Table 2. Suggestions concerning single group studies reporting in comparative effectiveness reviews

Report clearly in the Methods chapter if single group studies will be included
Report the rationale for including or excluding single group studies
If single group studies are included:
Report criteria used to select single group studies
Identify the explicit or implicit comparator (e.g., historical control, implied outcome based on known natural history)
Describe how single group studies are being used in the comparative effectiveness review to address Key Questions
Describe the methods used to assess quality/risk of bias and strength of evidence

Considerations When Using Single Group Studies in CERs

A second goal of this report is to illuminate the situations in which single group studies are commonly used in clinical investigation and review the considerations and assumptions required for their appropriate interpretation as a guide for the comparative effectiveness reviewer who encounters this type of evidence. We review these common single group study designs to aid in their interpretation by systematic reviewers and other readers of primary studies.

Although no direct, concurrent comparator is available in single group studies, both explicit and implicit comparisons are frequently made. Explicit comparisons are made when the investigators compare the single group of subjects before and after an intervention, or when the investigators choose to incorporate a historical comparator (e.g., historical data from the research institution or from an external cohort or existing database) in the analysis. Implicit comparisons are made when the expected outcomes in the absence of the intervention of interest are believed to be well known, and the expected effect size from the intervention is large. Depending on the single group study design, observed outcomes in the single treated group could be attributable to the intervention itself, a placebo effect, the natural course of the disease, or confounding by time-varying factors.[12] We discuss below these three common single group study designs and the key considerations and assumptions required for their interpretation (summarized in Table 3).

Table 3. Summary of challenges to interpretation of single group study designs

Study Design	Challenge
Before-after comparison of a single group of patients	Changes in factors other than the intervention of interest across the time periods compared are common (e.g., some patients decided to take daily fish oil supplements in the evaluation of newer lipid modifying agents)
	Natural recovery may occur (e.g., spontaneous resolution of acute sinusitis)
	Subject selection and attrition may be related to disease severity (e.g., some patients with hyperthyroidism dropped out of studies when they become euthyroid)
Comparison to historical control data	Changes in factors other than the intervention of interest across the time periods compared are common (e.g., changes in prevailing societal eating habits in different time periods in the evaluation of treatment for hypertension)
	Information may be unavailable or not reported on the variability of effect estimates from historical control groups
	Inadequate reporting of data sources for historical response rates may occur
Implicit comparison	Unpredictable or variable natural history of disease may occur (e.g., depression)

Differential (Before-After) Responses in a Single Group

These studies assess the difference in response before and after the administration of an intervention in a single group of patients. Patients serve as their own controls (before the intervention is administered).

The before-after design is commonly used in the evaluation of surgical interventions and other irreversible interventions (treatments or exposures). As an example, the change in blood pressure is assessed in patients with renal artery stenosis before and after vascular surgery (personal communication from K. Uhlig). The generally irreversible nature of the procedure ensures the permanence of the intervention, but the permanence of the intervention alone does not guarantee that observed post-intervention changes are attributable to the procedure.

Time-series analyses represent another example of a before-after single group study design, frequently strengthened with more data points. Time-series analyses follow a given group or region over time, typically without covariate data on individuals. As an example, a time-series study based on employment records merged with prescription records compared work days lost before and after a triptan became available.[13] Though time-invariant individual-level confounders do not limit this design, confounding can occur with time-varying factors such as secular trends in incidence rates of disease, clinical care, or cointerventions such as the addition of new services or other quality improvement interventions.

The interpretation of before-after single group studies for the purpose of informing CERs must include consideration of alternative explanations for observed treatment effects. For before-after treatment comparisons, it must be assumed that the intervention of interest presents the only change across the time periods compared if patient status at baseline is to represent what would have happened in the absence of treatment. The influence of adjunctive therapies administered concurrently, or carryover effects from therapies administered before the intervention of interest should be considered. Furthermore, natural recovery, the reduction or disappearance of symptoms regardless of the administration of a given treatment, presents another potential explanation for an observed before-after improvement in a health outcome in a single group comparison. Drawing valid and meaningful inferences about treatment effect using single group observational studies is problematic when evaluating conditions that are fluctuating or intermittent.[14] For example, spontaneous resolution of sinusitis is relatively common although antibacterial agents are often prescribed.[15] The potential for natural recovery diminishes with increasing disease duration and thus will most significantly affect studies selecting cases that have been recently diagnosed. On the other hand, single group studies of interventions for diseases with stable or steadily progressing courses may be useful.[6] For example, individuals with amyotrophic lateral sclerosis steadily decline in function over time, with spontaneous recovery virtually unknown.[16,17] A single group study of an intervention for these patients may be able to document meaningful treatment effects.

Single group before-after studies may also be compromised by issues related to subject selection and attrition. Patients selected into a study may represent a relatively extreme subset of the patient population with respect to disease severity and symptoms. For example, due to regression to the mean, a single group study that recruits patients with an incident diagnosis of hypertension may observe an improvement in blood pressure if patients sampled represented relatively severe cases at the baseline time point and there is natural variation in blood pressure over time. Similarly, if symptoms fluctuate over the course of disease, the exacerbation of symptoms may drive patient self-selection into a given study population. In this scenario, treatment effects observed in before-after single group studies may be attributable to the

selective sampling of patients at a peak severity in the natural history of disease that have a tendency to return to average severity levels over time regardless of interventions administered. Also, if patients with either more or less favorable outcomes are lost-to-followup, the observed effect of the intervention will be under- or overestimated, respectively. The site of the subject selection may also affect the outcomes of the single group study given patient differences across practice sites. For example, there may be variability in the stage of disease presentation of pathological features among cases recruited from an academic medical center versus community health centers.[18] Careful description of the study population is needed to draw meaningful and appropriate inferences.

Single Group Studies With Explicit Historical Comparisons

Single group studies may report comparative statistical or qualitative analyses comparing results with data obtained from the study with historical data from the research institution or from an external cohort or existing database not drawn from the same institution or population. An example is the comparison of one cohort from one time period that had received one technique for cardiac ablation to abolish atrial fibrillation with another cohort within the same institution from a different time period that had received a different technique.[19]

Such studies are often used to estimate the impact of systems-level interventions. For example, in an evidence report on the utility of screening bilirubin values for preventing chronic bilirubin encephalopathy,[20] three retrospective single group studies were identified that compared rates of phototherapy or readmission for hyperbilirubinemia before and after the implementation of a screening program.[21-23] These studies were deemed to be of low methodological quality, but represented the only available evidence on the effect of a systems-level implementation.

The interpretation of these single group studies can be complicated by specific challenges to the validity and precision of historical comparisons. The appropriateness of historical data to represent what would have occurred in the absence of the intervention is limited by patient temporal drift, a systematic, populationwide shift in outcomes that may be attributable to changes in disease staging, imaging, supportive care, evolving therapies, or other concomitant factors.[24] Changes in patient populations in newer versus older trials can also invalidate historical comparisons. Differences between the patient populations of a single group trial versus historical control trials can arise due to differences in accrual sites (academic medical center versus community health centers) or changes in patient characteristics such as age, performance status, or other prognostic factors. For example, more recently diagnosed patients may have milder manifestations of a condition due to improved (and therefore likely increased) diagnostic sensitivity. Treatment effects may also be attributable to secular trends in clinical care. For example, a decrease in diabetes-related preventable hospitalizations was observed in the U.S. from 1998 to 2006, possibly indicating improvements in quality of primary care or a higher threshold for hospitalization for individuals with diabetes.[25] Using a before-after single group design to evaluate the treatment effect of an antidiabetic agent introduced during this timeframe could be confounded by these examples of temporal drift. In addition, in phase 2 studies the variability of effect estimates from historical control groups is often overlooked and can result in erroneous conclusions regarding the effect of the experimental treatment.[26,27] Inadequate reporting of data sources further impedes appropriate interpretation of findings from single group studies with historical comparisons. Roughly half of phase 2 trials did not cite the source of their historical response rates.[28]

Several simulation studies have been conducted to illuminate the debate concerning the use of single group studies with historical controls versus randomized controlled trial designs in phase 2 studies. One such study concluded that if a small phase 2 study is planned with few patients available, then the single group study is usually adequate, assuming the historical control response rate is well-known.[29] Randomized trials were preferred for larger trials and in the setting of high variability in historical response rates.[29] A second study used simulations to compare the false positive (α) and false negative (β) rates in single group versus randomized two-arm trial designs in the setting of random and systematic variability in historical control data. This analysis indicated that variability in historical control response rates and changes in patient populations in newer versus older trials affects both false-positive and false-negative error rates in single group studies, but did not impact error rates in the randomized two arm trial designs (though the two arm studies required a 2–4 fold increased sample size).[24] A third simulation study compared optimal single group and randomized phase 2 designs with a variety of commonly used false positive (i.e. $\alpha = 0.05, 0.10$) and false negative (i.e. $\beta = 0.20, 0.10$) error rates. The objective of this simulation was to explore which phase 2 design was more likely to correctly identify active treatments so that they could proceed to phase 3 investigations. Simulation data indicated that conducting single group phase 2 studies would lead to a higher proportion of phase 3 trials being conducted using active agents when the true response rate for standard therapy was very low (<5 percent), in the presence of positive historical bias in estimating the null hypothesis (i.e., the estimate of the standard of care treatment effect is greater than the true effect), or when there was no historical bias and low variability in estimates of the historical control rate. On the other hand, randomized phase 2 trials performed better in the presence of a moderately active standard of care agent, and in the presence of negative historical bias (i.e., the estimate of the standard of care treatment effect is less than the true effect).[30] This study did not however account for the possibility of a new drug application submission immediately following phase 2.

Single Group Studies With Implicit Comparisons

Some single group studies do not make explicit comparisons to patient status before treatment or to an assembled historical control group, but rather make implicit comparisons based on knowledge about the natural history of the disease. In these instances, comparisons are extrapolations based on expected outcomes in the "missing" comparison group. Such comparisons are meaningful only when the expected outcomes in the absence of the intervention are well-known, and the expected effect size from the intervention is large. Investigations of anticancer agents for metastatic lung cancer present a useful example. The unlikelihood of spontaneous regression of this disease coupled with relatively impressive response rates as in the case of epidermal growth factor receptor (EGFR) inhibitor treatment of EGFR-mutated lung cancer, single group studies could be informative.[31,32] A second example is the use of antibiotics for bacterial meningitis—the natural course of the disease in the absence of antimicrobial therapy is so dramatic, there is little doubt that the beneficial outcomes are attributable to the intervention. However, for some diseases with an apparently uniformly poor prognosis, there may be subtle yet clinically relevant differences between patients who are enrolled in the single group study and those who don't qualify. Careful review of the study population and eligibility criteria is needed to make an assessment concerning external validity. For diseases with an unpredictable or variable natural history, a strong rationale must be available to justify reliance upon implicit comparison groups.

Single Group Studies and Subgroup Analyses

Subgroup analyses may be applied to single group studies to identify subsets of patients who may be more likely to experience adverse events or favorable health-related outcomes. In these designs, analyses may identify whether a variable is a significant independent predictor of a health outcome or adverse event while taking into account the separate influences of other predictors. However, a single group study cannot quantify treatment effect modification without an accompanying set of strong assumptions, since data on outcomes in the absence of treatment are excluded by design. For example, a single group study of a pharmaceutical agent that enrolls both men and women can estimate the effect of sex on the rate of an outcome of interest. However, evaluating whether the risk of an outcome event differs between men and women under a particular treatment more than it does in the absence of treatment (i.e., whether the relationship between the pharmaceutical agent and the outcome rate is modified by sex), is not possible in the absence of an untreated male group and an untreated female group. In order to estimate whether there is effect modification of treatment effects by sex in single group studies, assumptions would need to be made about the risk of outcome events among both untreated males and females.

Single Group Studies and Confounding

In the analysis of before-after responses in a single group, confounding by measured or unmeasured factors that are potentially time-varying across the time periods compared remains a critical threat. Statistical methods applied to the analysis of single group studies may be helpful to attempt to adjust for measured confounding factors, or to quantify the potential impact of unmeasured factors on effect estimates. Occupational and environmental studies that include only a single group of exposed individuals attempt to estimate standardized morbidity or mortality ratios by using age-, sex-, and race-specific rates from a general population in place of the unobserved counterfactual rates. The caveat to this approach is that there are likely to be other factors that differ between the exposed group and the reference population that will result in residual confounding of the standardized morbidity or mortality ratio. For time series studies, unmeasured confounders may be addressed by constructing a smooth function of time to serve a proxy for unmeasured confounders. Sensitivity analyses may be useful to generate hypotheses about the strength of an unmeasured confounder, such as a putative secular trend associated with improved outcomes in a before-after study, that would be required to produce an observed given treatment effect if in fact the intervention is not truly associated with any benefit.[33]

Conclusions

For many therapeutic strategies, the evidence base includes observational single group studies. Our review of published CERs to date indicates that single group studies are commonly included in EPC reports, especially in reporting of harms. However, the rationale for including these designs is infrequently reported and the methods relevant to their use are not clearly defined. As a step forward, we have developed recommendations to improve the reporting of single group studies. We have also reviewed commonly encountered single group study designs and highlighted the challenges in their interpretation and application to comparative questions as a guide for systematic reviewers.

Understanding the ideal comparison group for the effect of a treatment is essential to the fundamental definition of confounding and drawing causal inferences from all clinical research designs. Indeed, the selection of a comparison group that serves as a good proxy for the counterfactual outcome is a challenge even in analytic observational research designs with direct concurrent comparison groups. Guidance on the interpretation and use of observational research designs in general for CER has been developed.[34] Since single group designs are likely to continue to constitute part of the evidence base for certain interventions, it is important that a similar set of recommendations are developed to guide the use of single group studies in CERs.

References

1. Sackett DL, Rosenberg WM, Gray JA, et al. Evidence based medicine: what it is and what it isn't. 1996. Clin Orthop Relat Res. 2007 Feb;455:3-5. PMID: 17340682
2. Campbell D, Stanley J. Experimental and quasi-experimental designs for research. Chicago, IL: Rand-McNally; 1963.
3. Gill AL, Bell CN. Hyperbaric oxygen: its uses, mechanisms of action and outcomes. [Review]. QJM. 2004 Jul;97(7):385-95. PMID: 15208426
4. Janne PA, Gurubhagavatula S, Yeap BY, et al. Outcomes of patients with advanced non-small cell lung cancer treated with gefitinib (ZD1839, "Iressa") on an expanded access study. Lung Cancer. 2004 May;44(2):221-30. PMID: 15084387
5. Miller MA, Ulisney K, Baldwin JT. INTERMACS (Interagency Registry for Mechanically Assisted Circulatory Support): a new paradigm for translating registry data into clinical practice. J Am Coll Cardiol. 2010 Aug 24;56(9):738-40. PMID: 20723804
6. Agency for Healthcare Research and Quality. Methods Reference Guide for Effectiveness and Comparative Effectiveness Reviews. Rockville, MD: Agency for Healthcare Research and Quality. 2007;Version 1.0. 2007.
7. Hills M, Armitage P. The two-period cross-over clinical trial. Br J Clin Pharmacol.1979 Jul;8(1):7-20. PMID: 552299
8. Lathyris DN, Trikalinos TA, Ioannidis JP. Evidence from crossover trials: empirical evaluation and comparison against parallel arm trials. [Review]. Int J Epidemiol. 2007 Apr;36(2):422-30. PMID: 173011029.
9. Woods JR, Williams JG, Tavel M. The two-period crossover design in medical research. Ann Intern Med. 1989 Apr 1;110(7):560-66. PMID: 2923388
10. Bruening W, Launders J, Pinkney N, et al. Effectiveness of Noninvasive Diagnostic Tests for Breast Abnormalities [Internet]. Rockville, MD: Agency for Healthcare Research and Quality; 2006 Feb. Report No.: 06-EHC005-EF. PMID: 20704043.
11. Bruening W, Schoelles K, Treadwell J, et al. Comparative Effectiveness of Core-Needle and Open Surgical Biopsy for the Diagnosis of Breast Lesions [Internet]. Rockville, MD: Agency for Healthcare Research and Quality; 2012. PMID: 20704042.
12. Evans SR. Clinical trial structures. J Exp Stroke Transl Med. 2010 Feb 9;3(1):8-18. PMID: 21423788
13. Miller DW, Martin BC, Loo CM. Sumatriptan and lost productivity time: a time series analysis of diary data. Clin Ther. 1996 Nov;18(6):1263-75. PMID: 9001842
14. Glasziou P, Chalmers I, Rawlins M, et al. When are randomised trials unnecessary? Picking signal from noise. BMJ. 2007 Feb 17;334(7589):349-51. PMID: 17303884
15. Wasserfallen JB, Livio F, Zanetti G. Acute rhinosinusitis: a pharmacoeconomic review of antibacterial use. [Review]. Pharmacoeconomics. 2004;22(13):829-37. PMID: 15329029
16. Pastula DM, Coffman CJ, Allen KD, et al. Factors associated with survival in the National Registry of Veterans with ALS. Amyotroph Lateral Scler. 2009 Oct;10(5-6):332-38. PMID: 19922120
17. Testa D, Lovati R, Ferrarini M, et al. Survival of 793 patients with amyotrophic lateral sclerosis diagnosed over a 28-year period. Amyotroph Lateral Scler Other Motor Neuron Disord. 2004 Dec;5(4):208-12. PMID: 15799548
18. Iezzoni LI, Shwartz M, Moskowitz MA, et al. Illness severity and costs of admissions at teaching and nonteaching hospitals. JAMA. 1990 Sep 19;264(11):1426-31. PMID: 2391739
19. Jais P, Hocini M, Hsu LF, et al. Technique and results of linear ablation at the mitral isthmus. Circulation. 2004 Nov 9;110(19):2996-3002. PMID: 15520313
20. Ip S, Chung M, Trikalinos T, et al. Screening for Bilirubin Encephalopathy. Rockville, MD: Agency for Healthcare Research and Quality; 2009 Oct. PMID: 20722171.
21. Bhutani VK, Johnson LH, Schwoebel A, et al. A systems approach for neonatal hyperbilirubinemia in term and near-term newborns. JOGNN - J Obstet Gynecol Neonatal Nurs. 2006 Jul;35(4):444-55. PMID: 16881988

22. Eggert LD, Wiedmeier SE, Wilson J, et al. The effect of instituting a prehospital-discharge newborn bilirubin screening program in an 18-hospital health system. Pediatrics. 2006 May;117(5):e855-e862. PMID: 16651290

23. Petersen JR, Okorodudu AO, Mohammad AA, et al. Association of transcutaneous bilirubin testing in hospital with decreased readmission rate for hyperbilirubinemia. Clin Chem. 2005 Mar;51(3):540-44. PMID: 15738516

24. Tang H, Foster NR, Grothey A, et al. Comparison of error rates in single-arm versus randomized phase II cancer clinical trials. J Clin Oncol. 2010 Apr 10;28(11):1936-41. PMID: 20212253

25. Wang J, Imai K, Engelgau MM, et al. Secular trends in diabetes-related preventable hospitalizations in the United States, 1998-2006. Diabetes Care. 2009 Jul;32(7):1213-17.

26. Thall PF, Simon R. Incorporating historical control data in planning phase II clinical trials. Stat Med. 1990 Mar;9(3):215-28. PMID: 2188324

27. Williams DA. Extra-Binomial Variation in Logistic Linear Models. Journal of the Royal Statistical Society Series C (Applied Statistics) 1982;31(2):144-8.

28. Vickers AJ, Ballen V, Scher HI. Setting the bar in phase II trials: the use of historical data for determining "go/no go" decision for definitive phase III testing. Clin Cancer Res. 2007 Feb 1;13(3):972-76. PMID: 17277252

29. Taylor JM, Braun TM, Li Z. Comparing an experimental agent to a standard agent: relative merits of a one-arm or randomized two-arm Phase II design. Clin Trials. 2006;3(4):335-48. PMID: 17060208

30. Pond GR, Abbasi S. Quantitative evaluation of single-arm versus randomized phase II cancer clinical trials. Clin Trials. 2011 Jun;8(3):260-69. PMID: 21511687

31. Sordella R, Bell DW, Haber DA, et al. Gefitinib-sensitizing EGFR mutations in lung cancer activate anti-apoptotic pathways. Science. 2004 Aug 20;305(5687):1163-67. PMID: 15284455

32. Lynch TJ, Bell DW, Sordella R, et al. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. N Engl J Med. 2004 May 20;350(21):2129-39. PMID: 15118073

33. Greenland S, Lash TL. Bias analysis. In: Rothman KJ, Greenland S and Lash TL, eds. Modern Epidemiology. 3rd ed. Philadelphia: Lippincott–Williams–Wilkins; 2008. p. 345-80.

34. Norris S, Atkins D, Bruening W, et al. Selecting Observational Studies for Comparing Medical Interventions. In: Agency for Healthcare Research and Quality. Methods Guide for Comparative Effectiveness Reviews. Rockville, MD. 2008. PMID: 21433401. www.effectivehealthcare.ahrq.gov/ehc/products/196/454/MethodsGuideNorris_06042010.pdf

Acronyms

AHRQ	Agency for Healthcare Research and Quality
CER	Comparative Effectiveness Review
EGFR	Epidermal Growth Factor Receptor
EHC	Effective Health Care
EPC	Evidence-based Practice Center

Appendix A. Summary Table of Empirical Review Extractions

CER Name	Year	Pharmaceutical	Surgical	Medical Devices	Behavioral Intervention	Other	Included Single Group Studies for benefit?	Rationale for Inclusion, if Provided	Included Single Group Studies for Harms?	Rationale for Inclusion, if Provided	Total Number of Studies Included	Total Number of Single Group Studies Included	Terms Used
Gastroesophageal Reflux Disease	2005	Yes	Yes	Yes	No	No	Yes	To supplement data on long-term effect of surgery	Yes	Not stated	105	23	Cohort
Epoetin and Darbepoetin for Managing Anemia in Patients Undergoing Cancer Treatment	2006	Yes	No	No	No	No	Yes	Used to identify predictors of response to treatment	No	NA	103	Not clearly presented	prospective cohort studies
Safety of Analgesics for Osteoarthritis	2006	Yes	No	No	No	No	No	NA	Yes	RCT did not provide adequate information	351	Not clearly presented	Cohort
Off-Label Use of Atypical Antipsychotics	2007	Yes	No	No	No	No	No	NA	No	NA	84	NA	NA
Second-Generation Antidepressants in the Pharmacologic Treatment of Adult Depression	2007	Yes	No	No	No	No	No	NA	Yes	Not stated	200	Not clearly presented	Cohort study
Oral Diabetes Medications for Adults With Type 2 Diabetes	2007	Yes	No	No	No	No	No	NA	Yes	Not stated	216	Not clearly presented	Single group studies not explicitly called out

Comparative Effectiveness of Percutaneous Coronary Interventions and Coronary Artery Bypass Grafting for Coronary Artery Disease	2007	No	Yes	No	No	No	No	NA	No	NA	113	NA	NA
ACEIs and ARBs for Treating Essential Hypertension	2007	Yes	No	No	No	No	No	NA	No	NA	69	NA	NA
Drug Therapies for Rheumatoid or Psoriatic Arthritis	2007	Yes	No	No	No	No	Yes	Not stated	Yes	Not stated	156	Not clearly presented	
Treatments To Prevent Fractures in Men and Women With Low Bone Density or Osteoporosis	2007	Yes	No	No	Yes	No	No	NA	Yes	Not stated	566	31	Single group studies not explicitly called out
Therapies for Clinically Localized Prostate Cancer	2008	Yes	Yes	Yes	No	No	Yes	Few trials	Yes	Not stated	436	352	National data bases
Insulin Analogues in Premixed Formulations for Adults with Type 2 Diabetes	2008	Yes	No	No	No	No	No	NA	No	NA	45	NA	NA
Radiofrequency Catheter Ablation for Atrial Fibrillation	2008	Yes	No	Yes	No	No	No	NA	Yes	Not stated	120	100	Single group studies not explicitly called out
Lipid-Modifying Agents	2009	Yes	No	No	No	No	No	NA	No	NA	135	NA	NA
Medications To Reduce Risk of Primary Breast Cancer in Women	2009	Yes	No	No	No	No	No	NA	No	NA	123	NA	NA

ACEIs and ARBs Added to Standard Medical Therapy for Treating Stable Ischemic Heart Disease	2009	Yes	No	No	No	No	No	NA	No	NA	61	NA	NA
Radiotherapy Treatments for Head and Neck Cancer	2010	No	No	No	No	Radiotherapy	Yes	Used single arm studies for hypothesis generation	Yes	Used single arm studies for hypothesis generation	108	70	Single arm studies
In-Hospital Use of Recombinant Factor VIIa for Off-Label Indications vs. Usual Care	2010	Yes	No	No	No	No	No	NA	Yes	Rare harms	74	19	Noncomparative observational studies
Nonoperative and Operative Treatments for Rotator Cuff Tears	2010	Yes	Yes	No	Yes	Acupuncture	Yes	Not stated	Yes	Not stated	137	71	Single group studies not explicitly called out
Recombinant Human Growth Hormone in the Treatment of Patients With Cystic Fibrosis	2010	Yes	No	No	No	No	Yes	Not stated	Yes	Not stated	26	8	Single group studies not explicitly called out
Screening and Treatment of Subclinical Hypothyroidism or Hyperthyroidism	2011	Yes	No	No	No	No	No	NA	Yes	Not stated	8	2	Single group studies not explicitly called out
Traumatic Brain Injury and Depression	2011	Yes	No	No	Yes	Yes	Yes	Not stated	No	NA	2	1	Single group studies not explicitly called out

Therapies for Children With Autism Spectrum Disorders	2011	Yes	No	No	Yes	No	Yes	Lack of studies with comparison groups	Yes	Lack of studies with comparison groups	183	74	Studies without a comparison group; prospective case series; single crossover trial; single arm studies
DMARDs in Children With Juvenile Idiopathic Arthritis	2011	Yes	No	No	No	No	No	NA	Yes	To identify AEs that were not reported by other study designs	198	Not clearly presented	Case reports; Case series
Pain Management Interventions for Hip Fracture	2011	Yes	No	No	No	No	No	NA	No	NA	83	NA	NA
Early Diagnosis, Prevention, and Treatment of Clostridium difficile Infection	2011	Yes	No	No	No	No	No	NA	No	NA	29	NA	NA
Diagnosis and Treatment of Obstructive Sleep Apnea in Adults	2011	Yes	Yes	Yes	Yes	Yes	No	NA	Yes	Dearth of comparative surgical data	190	13	Noncomparative cohort studies
Nonpharmacologic Interventions for Treatment-Resistant Depression in Adults	2011	Yes	No	Yes	Yes	Yes	No	NA	No	NA	79	NA	NA
Terbutaline Pump for the Prevention of Preterm Birth	2011	Yes	No	Yes	No	No	No	NA	Yes	Limited availability of RCTs	15	2	Noncomparative studies; Observational studies

Antiepileptic Medications in Patients With Epilepsy	2011	Yes	No	No	No	No	Yes	Not stated	Yes	Not stated	68	9	Single group studies not explicitly called out
Adjunctive Devices for Patients With Acute Coronary Syndrome	2011	No	No	Yes	No	No	No	NA	No	NA	175	NA	NA
Off-Label Use of Atypical Antipsychotics – Update	2011	Yes	No	No	No	No	Yes	Limited availability of RCTs	Yes	Limited availability of RCTs	331	Not clearly presented	Case series; Open label; Chart review; Open label pilot study; Self-controlled case series
ADHD: Effectiveness of Treatment in At-Risk Preschoolers	2011	Yes	No	No	Yes	No	No	NA	No	NA	129	NA	NA