

Sept 2011

The purpose of this glossary is to define or describe terms used in comparative clinical effectiveness research for managed care and pharmacy professionals. For cost effectiveness terminology, readers are referred to other resources, such as *Health Care Cost, Quality, and Outcomes: ISPOR Book of Terms*, Lawrenceville, NJ: International Society for Pharmacoeconomics and Outcomes Research; 2003.

*Prepared by Lisa E. Hines, PharmD, Clinical Research Pharmacist, University of Arizona College of Pharmacy.
Review and editorial assistance provided by Mary Brown, PhD, Jason Hurwitz, PhD, Daniel C. Malone, RPh, PhD,
Ann Taylor, MPH, MCHES, and Terri L. Warholak, RPh, PhD, University of Arizona
and Elizabeth Sampsel, PharmD, MBA, BCPS, Academy of Managed Care Pharmacy*

A

A priori analysis: See planned analysis.

Absolute risk difference: See risk difference.

Absolute risk increase (ARI): See risk difference.

Absolute risk reduction (ARR): See risk difference.

Activities of daily living (ADL): See functional status.

Adherence: The consistency and accuracy with which a patient follows a recommended medical regimen.¹ Also called compliance. See also persistence.

Adverse effect: A harmful or undesirable outcome occurring during or after use of a drug or intervention where there is a reasonable possibility of a causal relation.²

Adverse event: A harmful or undesirable outcome occurring during or after use of a drug or intervention but not necessarily caused by it.²

Adverse reaction/ adverse drug reaction: An adverse effect specifically associated with a drug.²

Agency for Healthcare Research and Quality (AHRQ): The lead federal agency charged with improving the quality, safety, efficiency, and effectiveness of health care for all Americans.³ As one of 12 agencies within the Department of Health and Human Services, AHRQ supports health services research to improve health care quality and promote evidence-based decision-making. See also Effective Health Care Program. Website: <http://www.ahrq.gov>.

AHRQ: See Agency for Healthcare Research and Quality.

Alpha error: See Type I error.

Alternative hypothesis: The opposite of the null hypothesis.⁴ It is the conclusion when the null hypothesis is rejected.

Applicability: The extent to which the effects observed in published studies are likely to achieve similar results when the same intervention is applied to the population of interest under “real-world” conditions (i.e., typical practice). Also called external validity, generalizability.⁵

Association: A relationship between two variables (characteristics), such that as one changes, the other changes in a predictable way.⁶ A positive association occurs when one variable increases as another one increases. A negative association occurs when one variable increases as the other variable decreases.⁶ Association does not imply causation.⁷ Also called correlation.

Attrition: Loss of participants during the course of a study.⁶ Participants lost during the course of a study are often called dropouts. Also called lost to follow up.

Attrition bias: Systematic differences between comparison groups in withdrawals or exclusions of participants from the results of a study.⁶ For example, participants may drop out of a trial because of side effects of the intervention. Excluding these participants from the analysis could result in an overestimate of the effectiveness of the intervention or an underestimate of side effect rates, especially when the proportion dropping out varies by treatment group.

B

Bayes' theorem: A theorem used to update the probability of an event in the light of a piece of new evidence.⁶

Bayesian analysis: A statistical approach based on Bayes' theorem that can be used in single studies or meta-analysis.⁶ Bayesian analysis involves the use of existing and new information to estimate the risk that a person will experience an event.¹

Beta error: See Type II error.

Bias: A systematic error in study design or conduct that results in a distorted assessment of the intervention's impact on the measured outcomes.¹ In clinical trials, the main types of bias arise from systematic differences in study groups that are compared (selection bias), exposure to factors apart from the intervention of interest (performance bias), participant withdrawal or exclusion (attrition bias), or assessment of outcomes (detection

bias).⁶ Reviews of studies may also be particularly affected by reporting bias, where a biased subset of all the relevant data is available.⁶

Bias prevention: Aspects of study design or conduct intended to prevent bias. In clinical trials, such aspects include randomization, blinding, and concealment of allocation.⁶

Blinded study: An experimental study in which participants do not know the treatment they are receiving; investigators may also be blind to the specific treatments.⁴ Double blind means that neither participants nor investigators know which treatment the participants receive.⁴ However, the terms single blind, double blind and triple blind are not used consistently and are ambiguous unless those who are blinded are specified.⁶

Blinding: See blinded study.

C

Case-control study: An observational study that compares individuals with a specific disease or outcome of interest (cases) to individuals from the same population without that disease or outcome (controls) and seeks to find associations between the outcome and prior exposure to particular risk factors.⁶ This design is particularly useful where the outcome is rare and past exposure can be reliably measured. Case-control studies are usually retrospective, but not always.

Causality: An association between two characteristics that can be demonstrated to be due to cause and effect (i.e., a change in one causes change in the other).⁶ Experimental studies such as randomized controlled trials can be used to support causality.⁶ However, observational studies usually cannot determine causality.⁶ See the Bradford-Hill Criteria for assessing evidence of causation.^{8,9} Sometimes called causation or causal effect.

Centers for Education and Research on Therapeutics (CERTs): A national initiative to increase awareness of the benefits and harms of new, existing, or combined uses of therapeutics (drugs, medical devices, and biological products) through education and research.¹⁰ The CERTs program is a network of research centers, each focusing on a broad therapeutic theme. The program is funded

and run as a cooperative agreement by the Agency for Healthcare Research and Quality, in consultation with the U.S. Food and Drug Administration (FDA). Website: <http://certs.hhs.gov/>.

Clinical outcomes: Medical events occurring as a result of disease or treatment (e.g., stroke, disability, hospitalization). Also called clinical endpoint.¹

Clinical practice guideline: User-friendly, evidence-based, systematically developed statements to assist primary health providers and patients in making appropriate health care decisions.¹

Clinical trial: A prospective experimental study that tests the safety, efficacy, and/or effectiveness of a health care intervention intended to prevent, diagnosis, or treat a specific disease or condition in humans.¹ An umbrella term for a variety of designs of health care trials.⁶

Cluster randomized trial: A randomized controlled trial in which participants are randomly assigned to the intervention in groups (clusters) defined by a common feature, such as the same physician or health plan.¹¹

Cochran's Q test: The classical test used in meta-analysis to assess whether a set of individual studies is

heterogeneous.¹² An indication of the presence or absence of heterogeneity, but not the extent of heterogeneity. Also called Cochran's *Q* statistic. See also *I*² statistic.

Cohort: A group of participants who remain together in the same study over time.⁴

Cohort study: An observational study with a defined group of participants (the cohort) that is followed over time.⁶ Outcomes are compared between subsets of this cohort who were exposed or not exposed (or exposed at different levels) to a particular intervention or other factors of interest. A prospective cohort study identifies participants and follows them into the future. A retrospective (or historical) cohort study identifies participants from past records and follows them from a previous time point to the present.

Comparative effectiveness research (CER): The generation and synthesis of evidence that compares the benefits and harms of alternative methods to prevent, diagnose, treat, and monitor a clinical condition or to improve the delivery of care.¹³ The purpose of CER is to assist consumers, clinicians, purchasers, and policymakers to make informed decisions that will improve health care at both the individual and population levels.

Comparative effectiveness reviews: Systematic reviews that evaluate evidence on alternative interventions to help clinicians, policymakers, and patients make informed treatment decisions.¹⁴ Many comparative effectiveness reviews funded by the Agency for Healthcare Research and Quality's Effective Health Care (EHC) Program are developed by Evidence-based Practice Centers (EPCs).¹⁵ The other type of research review produced by the EHC Program is called a technical brief.¹⁵

Comparison group: See control group.

Compliance: See adherence.

Complications: A term often used to describe adverse events following surgery or other invasive interventions.²

Composite endpoint: Endpoints that capture the number of patients who experience one or more of several events of interest in clinical trials.¹⁶ Aggregates of individual endpoints may be used to increase the event rate and thus the statistical power of the study and to capture the overall impact of interventions.¹⁶ Study results with composite endpoints may be misleading if the individual endpoints are of varying clinical importance, the number of events in the more important components is small, or the magnitude of effect differs markedly across components.¹⁷ Also called composite outcome.

Concealment of allocation: The process used to ensure that the investigator enrolling a participant into a randomized controlled trial does not know the group to which the participant is assigned.⁶ This process is aimed at preventing selection bias and is distinct from blinding. Some attempts at concealing allocation are more prone to manipulation than others, and the method of allocation concealment is used as an assessment of the quality of a trial.

Confidence interval (CI): A measure of the uncertainty around the main finding of a statistical analysis.^{4, 6, 7} If the study were repeated multiple times, it is the range of values within which the mean for each trial would occur 95% of the time. Estimates of unknown quantities (e.g., odds ratio) are usually presented as a point estimate and a 95% confidence interval. Alternatives to 95%, such as 90% and 99%, are sometimes used. Wider intervals indicate lower precision and narrow intervals indicate greater precision.

Confidence limits: The upper and lower boundaries of a confidence interval.⁶

Confounder: See confounding variable.

Confounding variable: A variable (or characteristic) more likely to be present in one group of participants than another that is related to the outcome of interest and may potentially confuse (confound) the results.⁴ For example, if individuals in the experimental group of a controlled trial are younger than those in the control group, it will be difficult to determine whether a lower risk of death in one group is due to the intervention or the difference in ages (age is the confounding variable).⁶ Randomization is used to minimize imbalances in confounding variables between experimental and control groups.⁶ Confounding is a major concern in non-randomized studies.⁶ Also called confounder.

Consistency: The extent to which the effects from studies in a systematic review appear to have the same direction and magnitude.¹⁸ See also effect or effect size. Sometimes also refers to the reliability of a measure or study to reproduce the same results.

CONSORT: Acronym for Consolidated Standards of Reporting Trials.¹⁹ Encompasses various initiatives developed by the CONSORT Group to alleviate the problems arising from inadequate reporting of randomized controlled trials. Extensions of the CONSORT Statement have been developed for other types of study designs, interventions, and data.

Construct validity: The degree to which the items on a test or measurement scale actually represent the characteristic being measured (usually not observable).^{1,4} See also validity.

Content validity: The degree to which a test or measurement scale actually measures what it is designed to measure as determined by expert opinion.¹ For example, a depression scale that assesses only one symptom of depression (e.g., cries a lot) will have a low content validity. A depression scale that assesses all major symptoms of depression will have higher content validity. See also validity.

Control: 1. In a controlled trial, a participant in the group receiving placebo, no treatment, an active comparator, or standard of care) that serves as a comparator for the experimental intervention.^{4,6} Also called control participants. 2. In a case-control study, an individual in the group without the disease or outcome of interest.⁶ 3. In statistics, to adjust for, or take into account, extraneous influences or observations.⁶

Control group: Participants in the control arm of a study. See also control, controlled trial, experimental group, and treatment group.

Controlled trial: A clinical trial that has a control group.^{6,11} More specifically, an experimental study that compares the outcomes observed in one study group (or arm) receiving the intervention of interest (experimental group) to one or more comparison (control) group(s) receiving placebo, no treatment, an active comparator,

or standard of care. Such trials are not necessarily randomized. Also called controlled clinical trial.

Conventional treatment: See standard of care.

Correlation: See association.

Criterion validity: An indication of how well a test or scale predicts another related characteristic or outcome.^{1,4} May be tested when the results obtained by one instrument can be verified through an independent observation or another instrument that has already been validated, ideally a “gold standard” if one exists. See also validity.

Critical appraisal: The process of assessing and interpreting evidence by systematically considering its validity, results, and relevance.⁶

Cross-sectional study: An observational study that examines a characteristic (or set of characteristics) in a set of participants at a specific time or time period.^{4,11}

Cumulative meta-analysis: A meta-analysis that adds studies one at a time in a specified order (e.g. according to date of publication or quality) and the results are summarized as each new study is added.⁶ In a graph of a cumulative meta-analysis, each horizontal line represents the summary of the results as each study is added, rather than the results of a single study.

D

Data-derived analysis: See unplanned analysis.

Data dredging: Performing many analyses on the data from a study, for example looking for associations among many variables.⁶ The term is particularly used to refer to unplanned analyses, where there is no apparent hypothesis, and only statistically significant results are reported. Multiple statistical analyses on the same set of data increase the probability of making a Type I error (i.e., attributing a difference to an intervention when chance is a reasonable explanation).

Data mining: Data analysis techniques that use algorithms to detect patterns in large data sets containing numerous variables with unknown complex relations.²⁰

DEcIDE: See Developing Evidence to Inform Decisions about Effectiveness.

Detection bias: Systematic differences between comparison groups in how outcomes are ascertained, diagnosed, or verified.⁶ Also called ascertainment bias. See also bias.

Developing Evidence to Inform Decisions about Effectiveness (DEcIDE): The DEcIDE Network is a collection of research centers created by the Agency for Healthcare Research and Quality in 2005 to gather new knowledge and information on specific treatments and conduct studies on the outcomes, effectiveness, safety, and usefulness of medical treatments and services.²¹

Directness: The extent to which the evidence links the interventions directly to health outcomes.¹⁸ Indirect evidence can encompass surrogate outcomes or refers to situations when two or more bodies of evidence are needed to compare interventions.

E

Effect or effect size: A statistical estimate of the effect of an intervention or treatment in a study that is used to determine sample sizes, compare treatments, and combine results across studies in meta-analysis.²² Values can be negative or positive and have no units. Values greater than 0.8 are uncommon and indicate a significant impact of a treatment or intervention.

Effective Health Care (EHC) Program: Funds individual researchers, research centers, and academic organizations to work together with the Agency for Healthcare Research and Quality to produce effectiveness and comparative effectiveness research for clinicians, consumers, and policymakers.²³ The EHC Program: 1) reviews and synthesizes published and unpublished scientific evidence; 2) generates new scientific evidence and analytic tools; and 3) synthesizes research findings and/or generates and translates them into useful formats for various audiences. The EHC Program has three primary products: 1) research reviews: (comparative effectiveness reviews and technical briefs); 2) research reports; and 3) summary guides. Website: <http://www.effectivehealthcare.ahrq.gov/>.

Effectiveness: The extent to which an intervention works under real-world conditions (i.e., in practice).²⁴ Effectiveness studies involving drugs examine whether they work when they are used the way that most individuals take them. A treatment is effective when most individuals who have the disease would improve if they used the treatment. Effectiveness studies ask the question, “Does it work?” Clinical trials that assess effectiveness are sometimes called pragmatic trials, management trials, or practical trials.

Effectiveness review: Comprehensive reports based on available evidence that evaluate the effectiveness of interventions.¹⁵ They are similar to comparative effectiveness reviews except there may not be a clear comparator for interventions evaluated in effectiveness reviews. Evidence-based Practice Centers develop effectiveness reviews with funding through the Agency for Healthcare Research and Quality’s Effective Health

Double blind: See blinded study.

Care (EHC) Program. The other type of research review produced by the EHC Program is a technical brief.

Efficacy: The extent to which an intervention produces a beneficial result under ideal conditions (i.e., in clinical trials).^{6, 24} Efficacy trials ask the question, “Can it work?” Clinical trials that assess efficacy are called explanatory trials.

EHC Program: See Effective Health Care Program.

Eisenberg Center: The John M. Eisenberg Center for Clinical Decisions and Communications Science translates comparative effectiveness reviews and research reports created by the Agency for Healthcare Research and Quality’s Effective Health Care Program into short, easy-to-read guides and tools for use by consumers, clinicians, and policymakers.²⁵

Endpoint: See outcome.

EQUATOR: Acronym for Enhancing the QUALity and Transparency Of health Research.²⁶ The EQUATOR Network was launched to coordinate initiatives to promote transparent and accurate reporting of health research and to assist in the development of reporting guidelines. Website: <http://www.equator-network.org/>.

Estimate of effect: See treatment effect.

Evidence synthesis: The collation, combination, and summary of findings from a body of evidence.¹¹ Can be qualitative or quantitative (meta-analysis). See also systematic review and meta-analysis.

Evidence-based medicine: Conscientious, judicious use of current best scientific evidence in making decisions about patient care.²⁷

Evidence-based Practice Centers (EPCs): The Agency for Healthcare Research and Quality created the EPCs in 1997 to conduct research reviews for the Effective Health Care (EHC) Program.²⁸ The EPCs are located at medical schools, universities, or medical centers throughout the country.

The EPCs produce comparative effectiveness reviews or effectiveness reviews on medications, devices, and other health care services with the goal of helping patients, physicians, and policymakers make better decisions about treatments.

Experiment: See experimental study.

Experimental group: Participants in the experimental arm of a study that receive the intervention of interest. Also called treatment group.

Experimental intervention: An intervention under evaluation.⁶ In a controlled trial, an experimental intervention arm is compared with one or more control arms, and possibly with additional experimental intervention arms.

Experimental study: A study in which the investigators actively intervene to test a hypothesis.¹¹ It is called a trial or clinical trial when human participants are involved.⁴ See also controlled trial.

Explanatory trial: A controlled trial that seeks to measure the benefits of an intervention in an ideal setting (efficacy) by testing causal research hypotheses with the aim of understanding.^{29,30} Trials of health care interventions are often described as either explanatory or pragmatic. See also pragmatic trial.

External validity: The extent to which results provide a correct basis for generalizations to other circumstances (e.g., populations, settings).⁶ Also called generalizability, applicability. See also applicability.

F

Face validity: When an instrument appears to measure what it is intended to measure.¹ See also validity.

Failsafe N: A calculation used to account for publication bias that estimates the number of unpublished or unretrieved nonsignificant studies that would nullify or lower the significance in a meta-analysis.³¹ Also called file-drawer analysis.

False negative (FN): A test result that is negative for a person who has the disease.⁴

False positive (FP): A test result that is positive for a person who does not have the disease.⁴

Fixed-effect model: A model used in meta-analysis to calculate a pooled effect estimate using the assumption that all factors that could influence the effect size are the

same in all the studies, and therefore the true effect size is the same (fixed) in all studies.³² Since all studies share the same true effect, it follows that that observed effect size varies from one study to the next only because of the random error inherent in each study. An alternative model is the random-effects model.

Forest plot: A graphical representation of the individual results of studies included in a meta-analysis together with the combined meta-analysis result.⁶ The plot also allows readers to see the heterogeneity among the results of the studies.

Functional status: A measure of a person's ability to perform his or her daily activities, often called activities of daily living (ADL).⁴

G

Generalizability: See applicability and external validity.

GRACE Principles: Acronym for Good Research for Comparative Effectiveness Principles.^{33,34} An initiative to enhance the quality of observational comparative effectiveness research and to facilitate its use for decision-making about therapeutic alternatives.

Grey literature: Refers to information that is not published in easily accessible journals or databases.⁶ Examples include trial registries, conference abstracts, books, dissertations, monographs, and reports held by the Food and Drug Administration and other government agencies, academics, business, and industry.¹¹

H

Harm: The totality of all possible adverse consequences of an intervention.²

Hazard rate: The probability of an event occurring given that it has not occurred up to the current point in time.⁶

Hazard ratio: Represents the increased risk with which one group is likely to experience the outcome of interest.⁶ A measure of effect produced by a survival analysis. For example, if the hazard ratio for death for a treatment is 0.5, then treated patients are likely to die at half the rate of untreated patients.

Head-to-head trial: A controlled trial that compares two active treatments.¹¹

Health outcomes: Encompasses clinical, surrogate, and humanistic outcomes.³⁵ Examples include mortality, physiologic measures, clinical events, symptoms, functional measures, and patients' experience with care.

Health status: Functional capacity or a state of physiological and psychological functioning or well-being.¹

Health technology assessment (HTA): A form of policy research that examines short-and long-term consequences of application of a health care technology.¹ The goal of HTA is to provide policymakers with information on new treatments or interventions.

Health-related quality of life (HRQOL): A broad theoretical construct developed to explain and organize measures regarding evaluation of health status, attitudes, values, and perceived levels of satisfaction and general well-being related to either specific health conditions or life as a whole from the individual's perspective.¹ See also patient-reported outcomes.

Heterogeneity: A general term used to describe variation or diversity among studies or participants.^{6, 36, 37} Heterogeneity should be distinguished as clinical (differences between studies in key characteristics of the participants, interventions, or outcome measures),

methodological (differences in study design, conduct, and quality), or statistical (differences in reported effects). Statistical heterogeneity refers to the degree of variation in the effect estimates from a set of studies; it is also used to indicate the presence of variability among studies beyond the amount expected due solely to chance.

Heterogeneous: Describes a set of studies or participants with sizeable heterogeneity.⁶ The opposite of homogeneous.

Historical control: Previously collected observations used as control values against which treatment values are compared.^{4, 6} Risk of bias associated with historical controls relates to systematic differences between the comparison groups due to changes over time (e.g., in risks, prognosis, health care, etc.)

Homogeneous: 1. Similarity of participants, interventions, and measurement of outcomes across a set of studies.⁶ 2. In meta-analysis, used specifically to describe the effect estimates from a set of studies where they do not vary more than would be expected by chance.⁶

Humanistic intermediary: Factors that affect the formation of patients' opinions about the effects of disease or treatment on their lives and well-being (e.g., values, norms, perceptions).¹

Humanistic outcomes: Patient self-assessment of the impact of disease or treatment on their lives and well-being (e.g., satisfaction, quality of life).¹ See also patient-reported outcomes (PRO).

Hypothesis: A conjectural statement of the relation between two or more variables.³⁸ A proper hypothesis should be pre-specified, measurable, have theoretical or empirical support, be clearly articulated, and testable by an appropriately designed study.⁶ See also null hypothesis.

I

I^2 statistic: A statistical test used to quantify heterogeneity in a meta-analysis.³⁹ It describes the percentage of variability in effect estimates due to heterogeneity rather than sampling error (chance).⁶ Also called I^2 index. See also Cochran's Q test.

Incidence: The number of new cases of an event that develop within a given time period in a defined population at risk, expressed as a proportion.¹

Intention-to-treat (ITT) analysis: In a randomized controlled trial, the statistical analysis of all participants based on the group to which they were originally assigned.^{4,6} This minimizes bias caused by the loss of participants (attrition) that may disrupt the baseline equivalence established by randomization. The term is often misused in trial publications when some participants were excluded.

Intermediate outcome: See surrogate endpoint.

Internal consistency: See reliability.

Internal validity: The extent that the design and conduct of a study are likely to have prevented bias.⁶ More rigorously designed (better quality) trials are more likely to yield results that are closer to the truth. See also validity, bias prevention.

Intervention: A generic term used to describe a program, policy, measure or activity designed to have an impact on an illness or disease in an individual or a population.¹ In clinical trials, the term may be used to describe regimens in all comparison groups.⁶

J

John M. Eisenberg Center: See Eisenberg Center.

L

Level of significance: The probability of incorrectly rejecting the null hypothesis in a test of a hypothesis.⁴ See also p value.

Literature overview: A narrative summary of a specific topic.¹

Literature review: A narrative summary of existing published literature of a specific topic.¹

M

Meta-analysis: Use of statistical techniques in a systematic review to combine results from multiple individual studies.^{4,6,40} Encompasses a wide variety of methodological approaches whose goal is to quantitatively synthesize and summarize data across a set of studies. Typically, the objective of the analysis is to increase the precision and power of the overall estimated effect of an intervention by producing a single pooled

estimate. Sometimes misused as a synonym for systematic reviews. Also called quantitative synthesis.

Multiple comparisons: Performance of multiple analyses on the same data.⁶ Multiple statistical comparisons increase the probability of making a Type I error (i.e. attributing a difference to an intervention when chance is a reasonable explanation). See also data dredging.

N

Negative predictive value (NPV): The proportion of individuals with a negative test result who do not have the disease (true negative), and can be interpreted as the probability that a negative test result is correct.^{1,6} Calculation: $NPV = TN / (TN + FN)$, where TN = true negative and FN = false negative.

Negative study: A study that does not have “statistically significant” results.⁴¹ The term can generate confusion because it refers to both statistical significance and the direction of effect; studies often have multiple outcomes; the criteria for classifying studies as negative or positive are not always clear; and, in the case of studies of risk or undesirable effects, “negative” studies are ones that do not show a harmful effect.⁶

NNH: See number needed to harm.

NNT: See number needed to treat.

Nonexperimental study: See observational study.

Nonrandomized trial: A clinical trial in which subjects are assigned to treatments on other than a randomized basis.⁷

Nonsystematic error: Random error that is always present in measurement. Nonsystematic error can be estimated and reduced using statistical methods.⁴² See also random error.

Null hypothesis: The hypothesis being tested about a population, where null generally means “no difference” and thus refers to a hypothesis that no differences between groups or relationships between variables will be found.⁴

Number needed to harm (NNH): The average number of patients who need to be treated over a specific period of time to cause one additional undesirable outcome (or one fewer to experience a beneficial outcome) by the end of the period.^{1,4,6} It is the reciprocal of the absolute risk increase (ARI) or risk difference. Calculation: $NNT = 1 / ARI$. Also called number needed to treat to harm (NNTH).

Number needed to treat (NNT): The average number of patients who need to be treated over a specific period of time to promote one additional beneficial outcome (or prevent one additional undesirable outcome) by the end of the period.^{1,4,6} It is the reciprocal of the absolute risk reduction (ARR). Calculation: $NNT = 1 / ARR$. Also called number needed to treat to benefit (NNTB).

O

Observational study: A study in which investigators observe the course of events and do not assign participants to the intervention.^{6,11} Also called non-experimental study.

Odds: A way of expressing the chance of an event, calculated by dividing the number of individuals in a sample who experienced the event by the number for whom it did not occur.⁶ For example, if in a sample of 100, 20 individuals died and 80 individuals survived, the odds of death are $20 / 80 = 1/4$, 0.25 or 1:4.

Odds ratio (OR): An estimate of the relative risk calculated in case-control studies.^{4,6} It is the ratio of the odds of an event in one group to the odds of an event in

another group. In studies of treatment effect, the odds in the treatment group are usually divided by the odds in the control group. An odds ratio of 1 indicates no difference between comparison groups. For undesirable outcomes, an OR of < 1 indicates that the intervention was effective in reducing the risk of that outcome.

Outcome: The result of an experimental study that is used to assess the effect of an intervention.^{4,6} Also called endpoint.

Outcomes research: Evaluation of the effect of health care interventions on patient-related, clinical, humanistic, and economic outcomes.¹

P

P value: The probability (ranging from zero to one) that the results observed in a study (or more extreme results) could have occurred by chance if in reality the null hypothesis were true (refers to a Type I error).⁶

Patient registry: An organized system that uses observational study methods to collect uniform data (clinical or other) to evaluate specified outcomes for a population defined by a particular disease, condition, or exposure, and that serves one or more predetermined scientific, clinical, or policy purposes.⁴³

Patient-centered outcomes research (PCOR): Research focusing on the outcomes of concern to patients; these may include three major categories of patient-assessed health outcomes: 1) health status (encompassing health-related quality of life and functional status); 2) health utilities (patients' values for a particular state of health); and 3) patient satisfaction.⁴⁴

Patient-Centered Outcomes Research Institute (PCORI): An independent organization created to help patients, clinicians, purchasers and policy makers make better informed health decisions.⁴⁵ PCORI commissions research that reflects and supports patients' values and interests to provide reliable, evidence-based information for the health care choices patients and their caregivers they face.⁴⁵ The American Reinvestment and Recovery Act (ARRA) allocated \$1.1 billion for comparative effectiveness research (CER) and the Patient Protection and Affordable Care Act established PCORI to promote ongoing CER.⁴⁶

Patient-reported outcomes (PRO): An umbrella term that refers to outcome data reported directly by the patient.¹ This is one source of data that may be used to describe a patient's condition and response to treatment. It includes such outcomes as global impressions, functional status, well-being, symptoms, health-related quality of life, satisfaction with treatment, and treatment adherence.

Per protocol analysis: An analysis of the subset of participants from a randomized controlled trial who complied with the protocol sufficiently to ensure that their data would be likely to exhibit the effect of treatment.⁶ This subset may be defined after considering exposure to treatment, availability of measurements, and absence of major protocol violations. This analysis strategy may be subject to bias because the reasons for noncompliance may be related to treatment.

Performance bias: Systematic differences between intervention groups in care provided apart from the intervention being evaluated.⁶ For example, if participants know they are in the control group, they may be more likely to use other forms of care. Health care providers might behave differently if they are aware of a patient's assignment to a particular study group. Blinding of study participants and providers of care is used to protect against performance bias.

Persistence: The continued use of the prescribed pharmacotherapeutic regimen or other program. Also called treatment persistence.¹ See also adherence.

Pharmacoepidemiology: Study of the use, effects, and outcomes of drug treatment from an epidemiological (population) perspective.¹

Pharmacovigilance: The scientific field of collecting, analyzing, and interpreting postmarketing reports with the intention to generate detect, and/or validate signals for potential side effects from marketed products.¹ See also postmarketing surveillance, data mining, and patient registry.

PICOTS: Acronym for Population, Intervention, Comparator, Outcome, Timing, and Setting.⁴⁷ Parameters developed for formulating questions and locating primary studies for inclusion in systematic reviews. Also useful for evaluating applicability.

Planned analysis: Statistical analysis specified in a study protocol that is planned in advance of data collection (in contrast to unplanned analysis).⁶ Also called a priori analysis, pre-specified analysis.

Point estimate: A value (statistic) obtained from sample data this is used as the best estimate of what is true for the relevant population from which the sample is taken.⁴ ⁶ A point estimate is a measure of central tendency (e.g., mean, median, mode) that alone does not consider variability (e.g., standard deviation, standard error). Often used as a general term for results (e.g., risk difference, odds ratio, relative risk) obtained from a sample (a study or meta-analysis).

Population: The entire collection (group) of observations or participants that have something in common (e.g., age, disease) and to which conclusions are being inferred.^{4,6}

Positive predictive value (PPV): The proportion of individuals with a positive test result who have the disease, and can be interpreted as the probability that a positive test result is correct.^{1,6} Calculation: $PPV = (TP) / (TP + FP)$, where TP = true positive and FP = false positive.

Positive study: A study with statistically significant results, usually indicating a beneficial effect of the intervention being studied.^{6,41} The term can generate confusion because it refers to both statistical significance and the direction of effect; studies often have multiple outcomes; the criteria for classifying studies as negative or positive are not always clear; and, in the case of studies of risk or undesirable effects, “positive” studies are ones that show a harmful effect.⁶ See also negative study and publication bias.

Postmarketing surveillance: The practice of monitoring a drug or device after marketing and is a component of the science of pharmacovigilance.¹ The primary aim is to evaluate safety, including the risk for specific adverse effects or for potential differences in the drug’s safety profile in special populations or disease states. Approaches to monitor the safety of drugs include spontaneous reporting databases, prescription event monitoring, electronic health records, and patient registries.

Power: The probability of rejecting the null hypothesis when a specific alternative hypothesis is true. The power of a hypothesis test is one minus the probability of Type II error.⁶ In clinical trials, power is the probability that a trial will detect, as statistically significant, an intervention effect of a specified size. Studies with a given number of participants have more power to detect large effects than small effect. In general, power is set at 80% or greater when calculating sample size. Also called statistical power.

Pragmatic trial: A controlled clinical trial designed to measure the benefit of an intervention in normal practice (effectiveness) to help guide decisions between options for care.^{29,30} Trials of health care interventions are often described as either explanatory or pragmatic. See also explanatory trial.

Precision: 1. In statistics, precision is the degree of certainty surrounding an effect estimate for a given outcome.⁶ The greater the precision, the less the

measurement error. Confidence intervals around the estimate of effect from each study are one way of expressing precision, with a narrower confidence interval meaning more precision. 2. In trial searching, precision is the proportion of relevant articles identified by a search strategy expressed as a percentage of all articles (relevant and irrelevant) identified by that strategy. Highly sensitive strategies tend to have low levels of precision. Calculation: Precision = number of relevant articles / number of articles identified.

Predictive value: A measure of the usefulness of a screening/diagnostic test.^{1,6,48} The probability that an individual with a positive test is a true positive is referred to as the positive predictive value of a test. In contrast, the negative predictive value of a test is the probability that the individual with a negative test is a true negative. Predictive value is related to the sensitivity and specificity of the test and the prevalence of the disease in the population tested.

Prevalence: The proportion of a population that is affected by a given disease or condition at a specified point in time.⁴ It is not truly a rate, although it is often incorrectly called prevalence rate.

PRISMA: Acronym for Preferred Reporting Items for Systematic reviews and Meta-Analyses.^{48,49} PRISMA is the major reporting guideline for systematic reviews and meta-analyses. Website: <http://www.prisma-statement.org>.

Prospective study: In evaluations of the effects of health care interventions, a study in which participants are identified according to current risk status or exposure, and followed forward through time to observe outcomes.⁶ Randomized controlled trials are always prospective studies. Cohort studies are commonly either prospective or retrospective, whereas case-control studies are usually retrospective. See also retrospective study.

Publication bias: The tendency of research with positive (statistically significant) results to be submitted and published more than research with negative or neutral (null or non-significant) results.⁵⁰ Publication bias is a type of reporting bias. See also reporting bias.

Q

Quality: The extent to which all aspects of a study's design and conduct can be shown to protect against systematic and nonsystematic bias and inferential error.⁵¹ See also bias prevention and internal validity.

Quality of care: The degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge.⁴⁰

Quality score: A value assigned to represent the validity of a study either for a specific criterion or overall.⁶ Quality scores are expressed as letters (A, B, C) or numbers. See also bias prevention.

Quality-adjusted life year (QALY): A universal health outcome measure applicable to all individuals and all diseases that allows for comparisons across diseases and interventions.¹ One QALY is a year of life with no disability

(i.e., perfect health). Cumulated across multiple years, the QALY combines, in a single measure, gains or losses in both quantity of life (mortality) and quality of life (morbidity).

Quantitative synthesis: See meta-analysis.

Quasi-experiment: A study that is similar to a true experiment except that it lacks random assignment of participants to treatment and control groups.^{52, 53} A quasi-experimental design may be used to reveal a causal relationship in situations where the researcher is unable to control all factors that might affect the outcome. Because full experimental control is lacking, the researcher must thoroughly consider threats to validity and uncontrolled variables that may account for the results.

R

Random allocation: A method that uses chance to assign participants to comparison groups in a trial, e.g. by using a random numbers table or a computer-generated random sequence.⁶ Random allocation implies that each individual (or unit) entered into a trial has the same chance of receiving each of the possible interventions. Also called random assignment.

Random error: Variation in a sample that can be expected to occur by chance.^{4, 6} Confidence intervals and p values allow for the existence of random error, but not systematic errors (bias). Also called nonsystematic error, random variation.

Random sample: A sample of n participants (or objects) selected from a population so that each has an equal and independent chance of being selected for the sample.⁴ Distinct from randomization and random allocation.

Random-effects model: A statistical model used in meta-analysis that assumes the true effects are normally distributed.³² Both within-study sampling error (variance) and between-studies variation are included in the assessment of the uncertainty (confidence interval) of the results.⁶ When there is heterogeneity among the results of the included studies beyond chance, random-effects

models will provide wider confidence intervals than fixed-effect models.⁶ See also fixed-effect model.

Randomization: The process of randomly assigning participants to one of the arms of a controlled trial.⁶ Ensures that participants have an equal and independent chance of being in each arm of the study. There are two components to randomization: generation of a random sequence and its implementation, ideally in such a way that those enrolling participants into the study are not aware of the sequence (concealment of allocation).

Randomized controlled trial (RCT): An experimental study (controlled trial) in which participants are randomly assigned to treatment groups (experimental and control groups).^{4, 11}

Rate: The speed or frequency that an event occurs, usually expressed with respect to time.⁶ For example, a mortality rate may be the number of deaths per year, per 100,000 individuals.

Recall bias: Bias arising from errors in recollecting events due to failures of memory and looking at things "in hindsight," with possibly changed views.⁶ This bias is a threat to the validity of retrospective studies.

Reference population: The population to which the results of a study can be generalized. See also external validity.⁶

Regression analysis: A statistical modeling technique used to estimate or predict the influence of one or more independent variables on a dependent variable, e.g. the effect of age, sex, and educational level on the prevalence of a disease.⁶ Logistic regression and meta-regression are types of regression analysis.

Regression toward the mean: The phenomenon in which the results observed are influenced by a tendency for groups to reflect the grand population mean value.⁵² Regression to the mean is problematic when one group is selected on the basis of extreme values, and the comparison group is not. This is a common issue with disease-state management programs, which select outliers in one time period but “regress” to the mean value in subsequent time periods.

Relative risk (RR): See risk ratio.

Relative risk reduction (RRR): The proportional reduction in risk in one treatment group compared to another.⁶ Calculation: $RRR = 1 - \text{risk ratio}$, usually expressed as a percentage. For example, if the risk ratio is 0.25, then the relative risk reduction is $1 - 0.25 = 0.75$ or 75%.

Reliability: The extent to which an instrument, scale, or other type of measurement or procedure yields consistent and reproducible results.^{4,6,7} Reliability is context-specific rather than a property of an instrument under all conditions. Lack of reliability can arise from divergences between observers or measurement instruments, measurement error, or instability in the attribute being measured. See also consistency.

Reporting bias: A bias caused by only a subset of all the relevant data being available.⁶ Studies in which an intervention is not found to be effective are sometimes not published. Because of this, systematic reviews that fail to include unpublished studies may overestimate the true effect of an intervention. In addition, a published report might present a biased set of results (e.g. only outcomes or sub-groups where a statistically significant difference was found). See also publication bias.

Representative population (or sample): A population or sample that is similar in important ways to the population to which the findings of a study are generalized.⁴

Research report: A report of accelerated practical research studies about the outcomes, comparative clinical effectiveness, safety, and appropriateness of health care items and services; one of the products from the Agency for Healthcare Research and Quality’s Effective Health Care Program.¹⁵ The research is conducted by centers known as Developing Evidence to Inform Decisions about Effectiveness (DEClIDE) Centers, which are health research organizations with access to health information databases and the capacity to conduct rapid turnaround research. Also called new research report.

Research review: A comprehensive report based on available evidence (evidence synthesis) that evaluates benefits and harms of alternative interventions and indicates where more research is needed.¹⁵ The Agency for Healthcare Research and Quality’s Effective Health Care Program produces two types of research reviews: comparative effectiveness (or effectiveness) reviews and technical briefs.

Retrospective study: A study that looks backward in time at outcomes of interest that have already occurred before the study was initiated.⁶ Case-control studies are usually retrospective, cohort studies sometimes are, and randomized controlled trials never are. See also prospective study.

Risk: The proportion of participants experiencing the event of interest over a specified period of time.^{6,54} Often referred to as the event rate (experimental event rate and control event rate), however these terms confuse risk with rate. Calculation: $\text{Risk} = \text{number of events or newly affected persons} / \text{total persons observed}$, expressed as a proportion or a percentage. For example, if the event is observed in 25 out of 100 participants, the risk is 0.25 or 25%.

Risk difference: The difference in size of risk between two groups.⁶ For example, if one group has a 15% risk of contracting a particular disease, and the other has a 10% risk of getting the disease, the risk difference is 5%. Also called absolute risk difference, absolute risk reduction, or absolute risk increase, depending on the circumstances.

Risk factor: A term used to designate a characteristic that is more prevalent among participants who develop a given disease or outcome than among participants who do not.⁴

Risk ratio (RR): The ratio of risks in two groups. In intervention studies, it is the ratio of the risk in the experimental (exposed) group to the risk in the control (unexposed) group.⁶ A risk ratio of 1 indicates no difference in risk between the two groups. For undesirable outcomes, a risk ratio of < 1 indicates that the intervention was effective in reducing the risk of that

outcome (e.g., the event is less likely to occur in the experimental than control group). Risk ratio is calculated in cohort or prospective studies. Also called relative risk.

Robust: A term used to describe a statistical method if the outcome is not affected to a large extent by a violation of the assumptions of the method.⁴

S

Safety: Substantive evidence of an absence of harm.²

Sample: A subset of a population.⁴

Sampled population: The population from which the sample was taken.⁴

Selection bias: Systematic differences in groups that are compared (affects internal validity). Random allocation and adequate concealment of allocation protects against this bias.^{6, 11} Selection bias may also occur with systematic differences between those who are selected for study and those who are not (affects external validity). Selection bias may also apply to how studies are selected for inclusion in systematic reviews.

Sensitivity: 1) The proportion of time a diagnostic test is positive in individuals who have the disease or condition.^{1, 4, 6} A sensitive test has a low false-negative rate. Calculation: $\text{Sensitivity} = \text{TP} / (\text{TP} + \text{FN})$, where TP = true positive and FN = false negative. Also called true positive rate or detection rate. 2) In systematic review, sensitivity applies to article identification and is a measure of a search's ability to correctly identify relevant articles.⁶ Calculation: $\text{sensitivity} = \text{number of relevant articles identified by the search} / \text{total number of relevant articles from all searches}$. Also called recall.

Sensitivity analysis: An analysis used to determine how sensitive the results of a study or systematic review are to changes in how it was done.⁶ Sensitivity analysis is used to assess how robust the results are to uncertain decisions or assumptions about the data and the methods that were used.

Serious adverse event: Any adverse event with serious medical consequences, including death, hospital admission, prolonged hospitalization, and persistent or significant disability or incapacity.²

Severe adverse event: An adverse event that is severe (including "non-serious" adverse events).² For example, a rash could be "severe," but not "serious" (i.e., not resulting death, hospital admission, prolonged hospitalization, and persistent or significant disability or incapacity).

Side effects: Unintended drug effects (beneficial or harmful) when given at doses normally used for therapeutic effects.²

Single blind: See blind study.

Specificity: The proportion of time a diagnostic test is negative in individuals who do not have the disease or condition.^{1, 4, 6} A specific test has a low false-positive rate. Calculation: $\text{Specificity} = \text{TN} / (\text{TN} + \text{FP})$, where TN = true negative and FP = false positive.

Standard of care: The typical or usual treatment for a particular condition at that time.⁶ Also called conventional or standard treatment.

Strength of evidence: An evaluation of a body of evidence. Core domains include risk of bias, consistency, directness, and precision.⁵⁵ Strength of evidence grades for the Agency for Healthcare Research and Quality Effective Health Care Program's comparative effectiveness reviews are high, moderate, low, or insufficient. The high, medium, and low grades indicate the level of confidence that the evidence reflects the true effect. An insufficient grade indicates that evidence is unavailable or inconclusive.

STROBE: Acronym for Strengthening The Reporting of Observational Studies in Epidemiology.⁵⁶ The STROBE statement provides guidelines for reporting observational studies in epidemiology.

Summary guides: Plain-language guides for clinicians, consumers, or policymakers that summarize the findings of research reviews on the benefits and harms of different treatment options; one of the products from the Agency for Healthcare Research and Quality's Effective Health Care Program.¹⁵ The John M. Eisenberg Center translates comprehensive evidence reports (research reviews) into short guides (called consumer guides, clinician guides, and policymaker summaries).

Surrogate endpoint: Measurements of a patient's physical or biomedical status used as a surrogate for, or to infer the degree of, disease (e.g., blood pressure as a surrogate for stroke or heart attack).^{1, 6} Surrogate endpoints correlate with clinical outcomes but the relationship is not necessarily definitive. Also called clinical intermediary, intermediate outcome, or surrogate outcome.

Synthesis: See evidence synthesis.

Systematic error: Measurement error introduced into a study by its design, rather than due to random variation.^{4, 7} A systematic error is the same (or constant) over all observations. See also bias.

Systematic overview: See systematic review.

Systematic review: A structured literature review conducted in a systematic fashion using preset criteria and a protocol.¹ More specifically, it is a scientific investigation that focuses on a specific question and uses explicit, prespecified scientific methods to identify, select, assess, and summarize the findings of similar but separate studies (qualitative synthesis).¹¹ Systematic review is a type of evidence synthesis that may include quantitative synthesis (meta-analysis), depending on the available data.¹¹ See also evidence synthesis and meta-analysis.

T

Target population: The population to which the investigator wishes to generalize.⁴

Technical brief: A type of research review (evidence synthesis) intended to provide an early objective description of the state of science related to a new technology (clinical intervention or health care service) for which limited information exists to support definitive conclusions.⁵⁷ They also provide a possible framework to assess applications and implications of the intervention and describe ongoing research and future research needs.⁵⁷ Technical briefs funded by the Agency for Healthcare Research and Quality's Effective Health Care (EHC) Program are developed by Evidence-based Practice Centers.¹⁵ The other type of research review produced by the EHC Program are comparative effectiveness (or effectiveness) reviews.¹⁵

Tolerability: A patient's or participant's ability or willingness to tolerate or accept unpleasant drug-related adverse events without serious or permanent consequences.²

Toxicity: Refers to the quality of being poisonous (e.g., hepatotoxicity).²

Translational research: The process of applying discoveries generated from laboratory, clinical, or population studies into clinical applications.⁵⁸ There are two components to translational research: 1) The transfer of new understanding of disease mechanisms gained in the laboratory into the development of new methods for diagnosis, therapy, and prevention and their first testing in humans; and 2) The translation of results from clinical studies into everyday clinical practice and health decision making.⁵⁹ The ultimate aim is to ensure that new treatments and research knowledge actually reach the patients or populations for whom they are intended and are implemented correctly.⁵⁸

Treatment effect: The amount of change in a condition or symptom resulting from a treatment (compared to not receiving the treatment).⁶ It is commonly expressed as a risk ratio (relative risk), odds ratio, or risk difference. Also called estimate of effect.

Treatment group: See experimental group.

Treatment persistence: See persistence.

Trial: An experimental study involving humans, commonly called a clinical trial.⁴

Trim and fill method: A statistical method used to account for publication bias that adjusts a meta-analysis for the impact of missing studies.⁶⁰

True negative (TN): A test result that is negative in an individual who does not have disease.⁴

True positive (TP): A test result that is positive in an individual who has the disease.⁴

Type I error: The error that results if a true null hypothesis is rejected or if a difference is concluded when no difference exists.⁴ Also called alpha error, false alarm and false positive.

Type II error: The error that results if a false null hypothesis is not rejected or if a difference is not detected when a difference exists.⁴ Also called beta error, missed opportunity, and false negative.

U

Uncontrolled trial: A clinical trial that has no control group.⁶

Unplanned analysis: Statistical analysis that is not specified in the trial protocol and is generally suggested by the data.⁶ In contrast to planned analysis. Also called data-derived analysis, post hoc analysis.

V

Validation: The process of testing and accumulating evidence that supports the valid use or interpretation of results from a measure or study.¹

Validity: The degree to which a result (of a measurement or study) is likely to be true and free of bias (systematic errors).⁶ More broadly, the extent that accumulated evidence and theory is available to support the interpretation and use of results from a measure or study. See also internal validity, external validity. Types of

validity are usually accompanied by a qualifying word or phrase; for example, in the context of measurement, expressions such as construct validity, face validity, content validity, and criterion validity are used.

Variable: A characteristic of interest in a study that has different values for different participants or objects.⁴ For example, variables could be sex (male or female), weight, or test scores.

REFERENCES

1. Berger ML, Binglefors K, Hedblom EC, Pashos CL, Torrance GW. *Health Care Cost, Quality, and Outcomes: ISPOR Book of Terms*. Lawrenceville, NJ: International Society for Pharmacoeconomics and Outcomes Research; 2003.
2. Chou R, Aronson N, Atkins D. Chapter 7. Assessing harms when comparing medical interventions. In: *Methods guide for effectiveness and comparative effectiveness reviews*. AHRQ Publication No. 10(11)-EHC063-EF. March 2011; <http://www.effectivehealthcare.ahrq.gov>. Accessed 06/20/2011.
3. AHRQ Agency for Healthcare Research and Quality. <http://www.ahrq.gov/>. Accessed 06/20/2011.
4. Dawson B, Trapp RG. *Basic & Clinical Biostatistics. Fourth edition*. New York: Lange Medical Books/McGraw-Hill; 2004.
5. Atkins D, Chang S, Gartlehner G. Chapter 6. Assessing the applicability of studies when comparing medical interventions. In: *Methods guide for effectiveness and comparative effectiveness reviews*. AHRQ Publication No. 10(11)-EHC063-EF. March 2011; <http://www.effectivehealthcare.ahrq.gov>. Accessed 06/20/2011.
6. Glossary of terms in the Cochrane Collaboration. Version 4.2.5. Updated May 2005. 2005; <http://www.cochrane.org/glossary>. Accessed 06/20/2011.

7. Strom BL. *Pharmacoepidemiology*. Chichester: John Wiley & Sons; 1994.
8. Hill AB. The environment and disease: Association or causation? *Proc R Soc Med*. May 1965;58:295-300.
9. Phillips CV, Goodman KJ. The missed lessons of Sir Austin Bradford Hill. *Epidemiol Perspect Innov*. 2004;1:3.
10. Centers for Education & Research on Therapeutics (CERTs). <http://certs.hhs.gov/>. Accessed 06/20/2011.
11. IOM (Institute of Medicine). Finding what works in health care: standards for systematic reviews. Washington, DC: The National Academies Press; 2011.
12. Huedo-Medina TB, Sanchez-Meca J, Marin-Martinez F, Botella J. Assessing heterogeneity in meta-analysis: Q statistic or I² index? *Psychol Methods*. 2006;11:193-206.
13. IOM (Institute of Medicine). Initial national priorities for comparative effectiveness research. Washington, DC: The National Academies Press; 2009.
14. Lohr KN. Emerging methods in comparative effectiveness and safety: symposium overview and summary. *Med Care*. 2007;45(10 Supl 2):S5-8.
15. Effective Health Care Program. 2011; <http://www.effectivehealthcare.ahrq.gov/>. Accessed 06/10/2011.
16. Ferreira-Gonzalez I, Busse JW, Heels-Ansdell D, et al. Problems with use of composite end points in cardiovascular trials: systematic review of randomised controlled trials. *BMJ*. 2007;334:786.
17. Montori VM, Permyer-Miralda G, Ferreira-Gonzalez I, et al. Validity of composite end points in clinical trials. *BMJ*. 2005;330:594-596.
18. Norris S, Atkins D, Bruening. Chapter 4. Selecting observational studies for comparing medical interventions. In: *Methods guide for effectiveness and comparative effectiveness reviews*. AHRQ Publication No. 10(11)-EHC063-EF. March 2011; <http://www.effectivehealthcare.ahrq.gov/>. Accessed 06/20/2011.
19. CONSORT transparent reporting of trials. <http://www.consort-statement.org/>. Accessed 06/20/2011.
20. Fayyad UM, Piatetsky-Shapiro G, Smyth P. From data mining to knowledge discovery: An overview. In: Fayyad UM, Piatetsky-Shapiro G, Smyth P, Uthurusamy R, eds. *Advances in Knowledge Discovery and Data Mining*. Menlo Park, CA: MIT Press; 1996:1-34.
21. About the DEcIDE Network. <http://www.effectivehealthcare.ahrq.gov/index.cfm/who-is-involved-in-the-effective-health-care-program1/about-the-decide-network/>. Accessed 06/20/2011.
22. Light RJ, Pillemer DR. *Summing Up: The Science of Reviewing Research*. Cambridge, MA: Harvard University Press; 1984.
23. Effective Health Care Program. <http://www.effectivehealthcare.ahrq.gov/>. Accessed 06/10/2011.
24. Gartlehner G, Hansen RA, Nissman D, Lohr KN, Carey TS. A simple and valid tool distinguished efficacy from effectiveness studies. *J Clin Epidemiol*. 2006;59:1040-8.
25. About the Eisenberg Center. <http://www.effectivehealthcare.ahrq.gov/index.cfm/who-is-involved-in-the-effective-health-care-program1/about-the-eisenberg-center/>. Accessed 06/20/2011.
26. Welcome to the EQUATOR Network website: the resource centre for good reporting of health research studies. 2011; <http://www.equator-network.org/>. Accessed 06/09/2011.
27. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. *BMJ*. 1996;312:71-2.
28. About Evidence-based Practice Centers (EPCs). <http://www.effectivehealthcare.ahrq.gov/index.cfm/who-is-involved-in-the-effective-health-care-program1/about-evidence-based-practice-centers-epcs/>. Accessed 06/11/2011.
29. Schwartz D, Lellouch J. Explanatory and pragmatic attitudes in therapeutical trials. *J Chronic Dis*. 1967;20:637-48.
30. Zwarenstein M, Treweek S, Gagnier JJ, et al. Improving the reporting of pragmatic trials: an extension of the CONSORT statement. *BMJ*. 2008;337:a2390.
31. Becker BJ. Failsafe N or file-drawer number, in In: Rothstein HR, Sutton AJ, Borenstein M, eds. *Publication Bias in Meta-analysis: Prevention, Assessment and Adjustments*. Chichester, UK: John Wiley & Sons, Ltd; 2006.
32. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. *Introduction to Meta-analysis*. West Sussex, UK: John Wiley & Sons, Ltd; 2009.
33. Good ReseArch for Comparative Effectiveness. <http://www.graceprinciples.org/>. Accessed 06/20/2011.
34. Dreyer NA, Schneeweiss S, McNeil BJ, et al. GRACE principles: recognizing high-quality observational studies of comparative effectiveness. *Am J Manag Care*. 2010;16:467-71.
35. Clancy CM, Eisenberg JM. Outcomes research: measuring the end results of health care. *Science*. 1998;282:245-6.
36. EBM Glossary. <http://www.medicine.ox.ac.uk/bandolier/glossary.html>. Accessed 06/10/2011.

37. Fu F, Gartlehner G, Grant M, et al. Chapter 9. Conducting quantitative synthesis when comparing medical interventions. In: *Methods guide for effectiveness and comparative effectiveness reviews*. AHRQ Publication No. 10(11)-EHC063-EF. March 2011; <http://www.effectivehealthcare.ahrq.gov>. Accessed 06/20/2011.
38. Kerlinger FN. *Foundations of behavioral research*. Third ed. Fort Worth: Harcourt Brace Jovanovich College Publishers; 1986.
39. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ*. 2003;327:557-60.
40. IOM (Institute of Medicine). *Medicare: A strategy for quality assurance*. Washington, DC: National Academy Press; 1990.
41. Olson CM, Rennie D, Cook D, et al. Publication bias in editorial decision making. *JAMA*. 2002;287:2825-8.
42. Taylor JR. *An introduction to error analysis: The study of uncertainties in physical measurements*. 2nd ed. Sausalito, CA: University Science Books; 1997.
43. Gliklich RE, Dreyer NA. Registries for evaluating patient outcomes: A user's guide. AHRQ Publication No. 10-EHC049. 2nd ed. Rockville, MD: Effective Health Care Program; Agency for Healthcare Research and Quality; U.S. Department of Health and Human Services; September 2010.
44. Curtis JR, Martin DP, Martin TR. Patient-assessed health outcomes in chronic lung disease: what are they, how do they help us, and where do we go from here? *Am J Respir Crit Care Med*. 1997;156(4 Pt 1):1032-9.
45. Patient-Centered Outcomes Research Institute. <http://www.pcori.org/>.
46. Sox HC. Comparative effectiveness research: a progress report. *Ann Intern Med*. 2010;153:469-72.
47. Counsell C. Formulating questions and locating primary studies for inclusion in systematic reviews. *Ann Intern Med*. 1997;127:380-7.
48. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Ann Intern Med*. 2009;151:W65-94.
49. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med*. 2009;6:e1000097.
50. Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR. Publication bias in clinical research. *Lancet*. 1991;337:867-72.
51. Lohr KN, Carey TS. Assessing "best evidence": issues in grading the quality of studies for systematic reviews. *Jt Comm J Qual Improv*. 1999;25:470-9.
52. Campbell DT, Stanley JC. *Experimental and quasi-experimental designs for research*. Boston: Houghton Mifflin; 1963.
53. Shadish WR, Cook TD, Campbell DT. *Experimental and Quasi-experimental Designs for Generalized Causal Inference*. Boston: Houghton Mifflin; 2002.
54. Bjornson DC. Interpretation of drug risk and benefit: individual and population perspectives. *Ann Pharmacother*. Apr 2004;38(4):694-699.
55. Owens DK, Lohr KN, Atkins D. Chapter 10. Grading the strength of a body of evidence when comparing medical interventions. In: *Methods Guide for Effectiveness and Comparative Effectiveness Reviews*. AHRQ Publication No. 10(11)-EHC063-EF. March 2011; <http://www.effectivehealthcare.ahrq.gov>. Accessed 06/20/2011.
56. von Elm E, Altman DG, Egger M, Pocock SJ, Gotsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Ann Intern Med*. 2007;147:573-7.
57. Clancy CM, Slutsky J. Preface, in Williams JW, Coeytaux R, Wang A, Glower DD. Percutaneous heart valve replacement. Technical Brief No. 2. (Prepared by Duke Evidence-based Practice Center under Contract No. 290-02-0025.). August 2010; www.effectivehealthcare.ahrq.gov/reports/final.cfm. Accessed 06/20/2011.
58. Woolf SH. The meaning of translational research and why it matters. *JAMA*. 2008;299:211-3.
59. Sung NS, Crowley WF, Jr., Genel M, et al. Central challenges facing the national clinical research enterprise. *JAMA*. 2003;289:1278-87.
60. Duval S. The trim and fill method. In: Rothstein HR, Sutton AJ, Borenstein M, eds. *Publication Bias in Meta-analysis: Prevention, Assessment and Adjustments*. Chichester, UK: John Wiley & Sons, Ltd; 2006.