“Comparative effectiveness” research is the best tool available today for making decisions about which new medication, medical device, or diagnostic test is most supported by the evidence. The purpose of a systematic review is to synthesize the results of multiple primary studies using explicit and reproducible methods. Meta-analysis is a form of systematic review that goes one step further. Rather than qualitatively synthesizing the results of multiple studies, the purpose of a meta-analysis is to develop a quantitative summary of the evidence with the help of special statistical techniques. Whereas a systematic review may conclude that a certain drug is effective in preventing a morbid event, a meta-analysis will tell us that the drug is, say, 3.7 times more effective than a placebo in preventing that morbid outcome. Conceptually, the distinction between the 2 types of reviews is straightforward. Despite the appeal inherent in developing a quantitative synthesis, the decision to conduct, or to make decisions based on, a meta-analysis must be made carefully, with an understanding of the benefits and limitations of the method. This paper provides an overview of the meta-analysis process, highlights the strengths and weaknesses of the method, and offers guidance on how to interpret and judge the value of meta-analytic results.

Key Elements of Meta-Analysis

Meta-analysis is a method used to critically evaluate evidence in an attempt to develop a single synthesis of the results and, where appropriate, to use statistical methods to combine findings from different studies into a single “pooled estimate.” The pooled estimate is also called the “overall treatment effect,” even when...
referring to pooling of diagnostic tests, screening studies, or other areas that are not treatments per se. The reasons to perform a meta-analysis are to:

• Increase statistical power (relative to individual studies) and determine if a treatment effect exists by combining multiple trials
• Improve the precision of the measurement of a treatment effect
• Combine data from conflicting studies and determine if a treatment effect exists
• Explore the impact on the treatment effects associated with differences in the design, conduct, analysis, and results of individual studies.

A meta-analysis begins with a systematic review of the literature, followed by a statistical analysis. Generally, it proceeds according to these steps:
1. Formulation of a study question
2. Setting inclusion and exclusion criteria for studies to be reviewed
3. Searching the literature
4. Triaging articles
5. Retrieving data from the selected studies
6. Pooling the data by applying specific statistical techniques
7. Investigating sources of differences between studies
8. Summarizing and presenting results.

Selecting the Data
To conduct a meta-analysis, one must begin by formulating the study question. A well-focused study question has 4 components that are referred to as PICO: Population—a description of the patient population to be addressed
Intervention—a detailed description of the intervention or exposure to be investigated
Comparison—a well-defined comparison group
Outcome—a specific outcome.

For example, the following research question includes all 4 PICO elements. In adult patients with type 2 diabetes, does monotherapy with rosiglitazone improve A1c levels compared with metformin at 6 months? To minimize bias, inclusion and exclusion criteria should be identified before performing the literature search. Many of the inclusion criteria are determined as a direct result of the research question (eg, definition of the patient population or the outcome to be studied); other criteria (eg, study design) are determined by the research team.

Ideally, meta-analyses are limited to one type of study design. Significant problems are associated with meta-analyses that combine results across different study designs (eg, randomized controlled trials and case-control studies). Other inclusion and exclusion criteria may include years of publication, language of publication, and minimum study size. Because the quality of a meta-analysis is very dependent on the studies that comprise it, well-thought-out inclusion and exclusion criteria are necessary to ensure the validity of the conclusions.

Once selection criteria are established, an extensive literature search should be performed. A literature review typically begins with a systematic, well-documented, and repeatable search of electronic databases, such as PubMed, Embase, and/or the Cochrane Database of Systematic Reviews. The search should also include a “hand-search” of the reference lists of individual studies included in the meta-analysis. A well-designed search strategy will also utilize other sources, for example, “grey” literature (ie, not published in a peer-reviewed journal but in the form of reports, unpublished theses, books, and so on). A reliable literature search must be exhaustive and should be reproducible, with specific search terms reported in the methods section.

The literature search will produce an extensive list of potentially relevant citations. Articles must then be triaged for possible inclusion. The Quality of Reporting of Meta-Analyses (QUOROM) committee has created a flowchart that outlines a recommended format for presenting the “flow” of included and excluded studies. This chart clearly shows how the investigators move from the entire pool of articles identified to the final set of articles included in the pooled analysis. This allows readers to evaluate the number of and the rationale for all exclusions. Each study included in the meta-analy-
sis should be evaluated for quality. Ideally, this assessment should follow a generally accepted scale or checklist, such as the Jadad Score for randomized trials or STROBE (Strengthening the Reporting of Observational Studies in Epidemiology). Data such as measures of treatment effects, confidence intervals, study characteristics, and other variables hypothesized to affect the study results are then abstracted from the chosen studies, using a standardized form. The identification of studies and the data abstraction process should be carried out independently by at least 2 researchers (who resolve disagreements by consensus) to minimize bias in article selection and reduce errors in data collection.

### Statistical Analysis

The next step is to decide whether statistical pooling is justified. Although the goal is to obtain meaningful results from previous studies, there is also a possibility of reaching erroneous conclusions, especially if data are pooled from studies that are too diverse. If the differences between studies are significant—different patient populations or different comparison groups—combining results may not make clinical sense. In such cases, results should not be pooled but rather be presented in a systematic review. The extent of variability between studies is subjective, and meta-analyses are often criticized for inappropriately combining heterogeneous data (see below).

The results of a meta-analysis are graphically expressed in a type of graph called a “forest plot” (Figure 1), which displays the independent results of each study and the overall pooled result on the same plot. This allows the reader to quickly visualize the treatment effect in individual studies relative to the combined treatment effect across all studies. The results of each study are represented by squares that vary in size according to the relative size of the study (i.e., larger squares denote larger studies). Each square is bisected by a horizontal line, the length of which spans the 95% confidence interval for the treatment effect reported in that study. The overall pooled estimate is shown at the bottom of the figure, represented by a diamond, whose center and edges correspond to the mean overall result and confidence interval, respectively.

### How to Identify Potential Problems

Once the studies to be included in the pooled analysis have been identified, some differences will still be evident between studies. The nature and magnitude of these differences play a critical role in determining the...
methods to be used in the statistical analysis. Researchers have developed summary measures that quantify the degree of interstudy “heterogeneity.” Cochran’s Q (based on the chi-square) test is the most common test for heterogeneity.

If “significant” heterogeneity is identified (ie, \( P < .20 \) is a typical cut-off in meta-analysis), it is important to determine whether it is valid to combine heterogeneous studies into a single pooled measure of treatment effect. For example, in a meta-analysis of hormone replacement therapy (HRT) and breast cancer, the Q test identified heterogeneity (Figure 2). One study clearly stood out—the Million Women Study, not only because it was the largest and only study conducted in Europe but also because of other important underlying clinical differences. When this study was excluded, the Q test was no longer significant, and the point estimate better reflected the pooling of similar studies.\(^1\)

The \( I^2 \) statistic (also referred to as “inconsistency”) is an updated version of the Q test. This updated measure is designed to allow comparison of heterogeneity between meta-analyses with different numbers of pooled studies, which cannot be done with the Q test alone.\(^2\) An \( I^2 \) value of less than 25% is considered good, 25% to 50% is acceptable, and more than 50% is unacceptable.

When heterogeneity is identified by the Q test or by \( I^2 \), several approaches can be considered to address the differences. The first approach is to simply ignore the interstudy differences. Some investigators have argued that “one true effect” must underlie all studies on a given topic, and results should be pooled by use of a “fixed effects” model. A second method is to use a “random effects” model, which statistically accounts for heterogeneity and considers several “true effects.” (The difference between fixed and random effects models\(^3\) is beyond the scope of this review.) In general, a random effects model results in the same pooled treatment estimate but with a wider confidence interval than the fixed effects model. The majority of meta-analyses now report random effects results, and those that report only fixed effects results should be viewed with skepticism, especially those that report pooled treatment effects with marginal statistical significance.\(^4\)

Another common approach to addressing heterogeneity is using meta-regression, which involves applying traditional regression techniques—and their inherent explanatory power—to meta-analysis. In meta-regression, the dependent variable is the outcome (eg, the odds ratio or the relative risk), and the independent variables are study factors, such as publication year, country, study size, or type of drug, that may reasonably be expected to contribute to interstudy differences. Stratifying and conducting subgroup meta-analyses based on these factors may help identify important sources of heterogeneity (eg, the HRT example). Although meta-regression is useful to assess heterogeneity, the results should be considered “hypothesis generating.”

The final step in meta-analysis is to look for evidence of publication bias—the exclusion of negative studies that might have influenced the conclusions. Studies published in peer-reviewed journals tend to report positive results more than negative results. In practice, researchers who find negative results often do
It is important

In the absence of publication bias, the resulting
Cook DJ, Mulrow CD, Haynes RB. Systematic reviews: synthesis of
The authors searched the relevant
14
Different researchers may use different
Shah NR. What is the best evidence for making clinical decisions?
and that the longer it has been since pub-
15
Hopewell S, McDonald S, Clarke M, Egger M. Grey literature in
Figure 3. If publication bias is present, smaller studies
that fail to show a significant effect will likely be miss-
Figure 3 would be absent). Techniques
thus it’s best to confirm that publication bias assess-
Figure 3
The Limits of Meta-Analysis
Meta-analysis can be a powerful technique for sum-
Each meta-analysis is, in and of
itself, a scientific investigation, and its quality is
dependent on the methods used in carrying out the
“experiment.” Different researchers may use different
techniques, include different studies, and draw differ-
et conclusions. Like any experiment, meta-analyses
are subject to bias and error, both of which may affect
the validity of the conclusions and their utility for deci-
sion makers. As a result, not all meta-analyses are of
equal quality. Thus “consumers” of meta-analyses—
especially decision makers—must carefully assess the
quality of each meta-analysis by considering the
research questions asked, the methods used, the analy-
sis and interpretation of the data, the investigation of
heterogeneity, and the conclusions drawn.
Evaluating Quality
Several instruments for assessing the quality of a sys-
tematic review have been developed. It is important
to differentiate between the quality of the reporting of
a meta-analysis and the quality of the meta-analysis
itself. The report may, for a number of reasons (eg,
space limits, author preferences), omit important infor-
mation. The QUOROM statement offers a checklist of
evidence-based standards for reporting the results of
meta-analyses of randomized trials. The QUOROM
checklist—which identifies 18 key items to be includ-
ed in a report and explains how to describe them—is
widely used in the reporting and the evaluation of
published meta-analyses. The Meta-analyses of
Observational Studies in Epidemiology (MOOSE) is a
similar checklist that is used to guide the reporting of
meta-analyses of observational studies.
Over the past 20 years there has been tremendous
growth in the number of systematic reviews and meta-
analyses. A recent study investigated the average
length of time until a published systematic review
requires updating. The authors searched the relevant
literature to determine whether new evidence had
been published that would either invalidate the results
of a previous systematic review or would affect clinical
decision-making in an important way. In almost 60%
of the systematic reviews, the evidence showed that
the previously published review required updating.
The average length of time from publication to the
emergence of new evidence was about 5.5 years, with
25% of studies becoming outdated by 2 years. This sug-
gests that meta-analyses may have a relatively short
“shelf-life,” and that the longer it has been since pub-
lication, the more important it is to assess whether
new evidence has emerged that may change that
study’s conclusions.
Conclusions
Meta-analysis is a sophisticated tool for decision
makers. As with all medical evidence, however, sys-
tematic reviews and meta-analyses should be regarded
with due skepticism and be read critically, focusing on
the elements discussed here. As meta-analyses begin to
address the development of pooled estimates of cost,
adverse effects, and comparative effectiveness (in addi-
tion to efficacy and effectiveness), their relevance to
our understanding of how to define and measure value
in healthcare will continue to grow.

Disclosure Statement
Dr Shah has received unrestricted research grants
from GlaxoSmithKline, Novartis, AstraZeneca, Roche,
Berlex, and Pfizer. Mr Jones and Dr Blecker have noth-
ing to disclose.

References
2. Cook DJ, Mulrow CD, Haynes RB. Systematic reviews: synthesis of
1;126(5):376-380.
3. Greenhalgh T. Papers that summarise other papers (systematic
4. Hopewell S, McDonald S, Clarke M, Egger M. Grey literature in
meta-analyses of randomized trials of health care interventions.
AHDB Stakeholder Perspective

Meta-Analyses and Patient Care

With the proliferation of medical research studies over the past few decades, healthcare providers needed evidence synthesis more sophisticated than the traditional narrative review. Systematic reviews and meta-analyses have filled that void and are now routinely used by those on the front lines of patient care.

PAYORS: Payors have been slower to adopt these techniques, but have now been embracing them more fully. The broad and rigorous overview of a topic provided by such a synthesis, when conducted by independent and transparent entities, makes available the data they have long needed, but to which they have not had access. At the same time, payors remain cautious that these evidence summaries may include populations different from those in their managed population.

MANUFACTURERS: Manufacturers have been among the most refined in their use of meta-analyses, using them not only to get early signals of effectiveness but also to explore new possibilities in the lifecycle of drugs and devices through meta-regression and other advanced techniques. Although certainly considered an opportunity, meta-analysis also presents risk. For example, the Avandia meta-analysis led to the convening of an FDA Task Force to more closely review all data available on thiazolidinediones, which changed the labeling on this medication.

GOVERNMENT: The National Institutes of Health (NIH) and the Centers for Medicare & Medicaid Services (CMS) have only recently considered systematic reviews and meta-analyses as credible sources of evidence. Through funding “evidence-based practice centers” and comparative effectiveness reviews, the NIH and CMS have quickly become major sponsors of high-quality systematic reviews. In addition, the FDA is looking for new data sources and tools to efficiently monitor treatment effectiveness and safety, and meta-analyses provide a valuable source of fresh information after medications have been commercially available.

PATIENTS: With the increasing prevalence of so-called consumer-driven healthcare, consumers play an important role in assessing the quality and cost-benefit of various treatment options. Meta-analyses and systematic reviews can be an effective tool for consumers, but it is important to educate consumers about the strengths and limitations of the technique, and the applicability of the results to their specific decision or concern.

Scott R. Taylor, RPh, MBA
Associate Director, Industry Relations
Geisinger Health System