National Pharmaceutical Council

DEMYSTIFYING COMPARATIVE EFFECTIVENESS RESEARCH:
A CASE STUDY LEARNING GUIDE

ROBERT W. DUBOIS, MD, PhD
SYLVIA L. KINDERMANN, MPH

November 2009

A WHITE PAPER PREPARED FOR NPC BY

CERNER LifeSciences

AUTHORS:

Robert W. Dubois, MD, PhD
Chief Medical Officer, Cerner LifeSciences

Sylvia L. Kindermann, MPH
Sr. Research Associate, Cerner LifeSciences

Acknowledgments: The authors would like to acknowledge Diana Enyedi, Jennifer Graff, Andrea Hofelich, Les Paul, Gary Persinger, Sabrina Siddiqui, and Kimberly Westrich for their thoughtful contributions to this report.

Funded and published by the National Pharmaceutical Council.

For copies of this report, please contact the National Pharmaceutical Council at 703-620-6390 or visit www.npcnow.org.

National Pharmaceutical Council

TABLE OF CONTENTS

EXECUTIVE SUMMARY

I. Introduction

Decisions made by patients and physicians about the care that the patient will receive frequently involve choices among available therapies. Commonly, there are few research studies that compare alternative treatments and provide evidence to guide a particular decision. Even less plentiful are studies that compare treatment options in real-world settings. As the realization of this need for enhanced comparative information to achieve best outcomes has grown, attention has focused on comparative effectiveness research. The Institute of Medicine defines this Comparative Effectiveness Research (CER) as:

> "...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. 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."[1]

With the current national focus on improving our health care system, CER has received substantial attention. In addition, in early 2009, the American Recovery and Reinvestment Act allocated $1.1 billion in new funds to increase the number of new CER studies. Three important types of comparative effectiveness research include: 1) Randomized Controlled Trials (RCTs), 2) Meta-Analysis, and 3) Observational Studies. If conducted and interpreted correctly, these types of research can help to inform health care decision-making. If, however, such studies are conducted or interpreted incorrectly, the comparative answers from these studies may be inaccurate, or worse, misleading. This white paper will review the three CER study types, discuss their proper methodologic approaches, and provide tools for CER consumers to use when they review them. Although additional types of biomedical research are very useful in understanding disease and developing new drugs, the focus of this paper is solely on the most common types of research used for CER.

II. The Study Designs

Randomized Controlled Trials

RCTs are experiments wherein subjects in a population are randomly allocated into study and control groups to receive or not to receive an experimental intervention such as a new therapeutic biopharmaceutical or medical device. Although some RCTs include large numbers of patients to address broad population questions about competing therapies, studies of this type often apply the intervention to a carefully selected population while reducing the impact of external factors (such as comorbidities) that could also influence the outcome of interest. The controls required in such a study type sometimes impact whether

the findings may apply to a broader population. Even so, the strengths associated with RCT methodology, particularly in its lowered risk of bias and confounding, have allowed this type of study design to gain sufficient credibility that many now consider it to be the "gold standard" of clinical research.

Randomized Controlled Trials: Tips for the CER Consumer

- **RCTs can determine whether an intervention provides benefit in a very controlled environment.**
- **The controlled nature of an RCT may limit its generalizability to a broader population.**
- **Targeted therapy illuminated by carefully thought out subgroup analyses can improve the efficacious and safe use of an intervention.**

Meta-Analysis

Meta-analyses synthesize data across a series of similar studies, generally RCTs. This study design requires careful selection of which studies to include and which statistical methods to employ. Publication bias, which occurs when positive studies are more abundantly available on a topic due to negative studies not being published, can significantly impair the validity of this study type. However, a balanced meta-analysis can produce a summary estimate of the medical literature that underpins benchmarks of clinical decision-making.

Meta-Analysis: Tips for the CER Consumer

- **The results of a meta-analysis are highly dependent on the studies included (and excluded). Are these criteria properly defined and relevant to the purposes of the meta-analysis? Were the combined studies sufficiently similar?**
- **The statistical methodology can impact study results. Have there been reviews critiquing the methods used in the meta-analysis?**
- **Nothing is permanent. Emerging data may change the playing field, and meta-analysis results are only as good as the data and statistics from which they are derived.**

Observational Studies

Observational studies follow participants over a period of time to examine the potential associations between patients' exposure to treatment and health outcomes. These studies can be performed prospectively, observing patients in real time, or they can be retrospective analyses of existing databases. Given the abundance of billing or administrative databases,

comparisons of therapies and their outcomes will frequently use these databases, as they embody the experience of patients in real-world settings. While results from observational studies tend to generate hypotheses and require the more rigorous RCT for confirmation, long-term observational studies have provided substantial medical knowledge (e.g., demonstrating the link between elevated blood cholesterol and the development of heart disease).

Observational Studies: Tips for the CER Consumer

- **Observational studies provide an understanding of real-world care and its impact, but can be biased due to uncontrolled factors.**
- **Before accepting the findings from an observational study, consider whether confounding factors may have influenced the results.**
- **Observational studies can identify associations but cannot prove cause-and-effect relationships.**
- **The Grace Principles (www.graceprinciples.org) are an excellent source of further information about observational studies and their application in CER.**

III. Framework for Reviewing a New CER Report

Each of the CER research designs can answer certain questions, although each has limitations. Notwithstanding the differences among these designs, three general questions can be asked to determine the applicability of the research to your needs.

STEP 1: Consider for whom the findings are applicable.

Typically, the assessment of new research begins with a careful consideration of the quality of the study, or its internal validity. Alternatively, first assessing its external validity by relating the study to your own needs may be the most efficient course of action when interpreting findings of CER. If the study used a population, intervention, or outcomes that are not relevant to your own situation, there is little need to continue further. For example, a study investigating the effect of the H1N1 vaccine in a school-age population may not help an internist seeking evidence to inform the care of older individuals.

STEP 2: Consider whether any aspects of the study design might affect the results.

If the study has relevance (passes Step 1), then closely examine the study's validity. To do so, first consider whether the researchers selected a study design that matches the question being asked. For example, why did the authors conduct another RCT when a more real-world environment would have better answered the research question? Will the population in an observational study likely have so much patient heterogeneity that confounding cannot be avoided? As an

example, in routine care, physicians may prescribe a new antibiotic only for their sickest patients. In that circumstance, the antibiotic's real benefit may be masked since those sicker patients would ordinarily do worse than the healthier patients prescribed the alternative treatment. Are meta-analyses being conducted just because studies are available for synthesis or do the available studies truly address the question asked? After assessing the choice of study design, next examine whether that design was carried out with adequate rigor. To make that assessment, the issues will differ depending upon the methodology used in the CER.

STEP 3: Consider whether the findings may change with new research.

After completing these two steps (i.e., relevance and rigor), the reader should consider whether the apparently definitive conclusion may change over time. The inherent nature of research is cumulative, wherein new findings constantly emerge to build on older results. As such, all findings should be treated with that knowledge and with the expectation that further results may appear that expand upon, clarify, or reject these findings. Keeping the patient in mind, perhaps adopt the mindset: "How sure are we, really?" To answer this question, consider whether the area of research changes rapidly. If so, it raises the likelihood that advances in scientific knowledge will alter the current findings. For example, an increased understanding of the biology and genetics of cancer will probably influence the true benefits and harms in an ever-evolving target population.

Consider also the influence of the CER upon clinical policy. If the findings will profoundly affect access to a therapy or promulgated standards of care, greater confidence in the stability of the findings would be important. If the area of medicine has relative stability and the impact of the research has a more modest purview, the reader may accept less certainty.

Conclusion

The increase in comparative effectiveness research will present the reader with an exciting array of new data to help inform health care decisions. While CER provides information, the informed reader should examine it carefully and apply the findings with caution. As solid as those findings may appear, all studies have limitations. Important criticisms may arise, further analyses may refine the results, and the study may not be adequately generalizable to your environment. The goal of new public investment in CER is to improve health outcomes and to provide additional information to the medical community so that it will better understand what works and what does not work in clinical care. The informed consumer will have multiple sources of information from which to choose for evaluating myriad health options. Education as to how to evaluate this information will be critical if it is to be used effectively to achieve the goals described.

CHAPTER 1: COMPARATIVE EFFECTIVENESS RESEARCH

Introduction

Decisions made by patients and physicians about the care that the patient will receive frequently involve choices among available therapies. Commonly, there are few research studies that compare alternative treatments and provide evidence to guide a particular decision. Even less plentiful are studies that compare treatment options in real-world settings. As the realization of this need for enhanced comparative information to achieve best outcomes has grown, attention has focused on comparative effectiveness research. The Institute of Medicine defines this Comparative Effectiveness Research (CER) as:

> "...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. 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."[1]

Three important types of comparative effectiveness research include: 1) Randomized Controlled Trials (RCTs), 2) Meta-Analysis, and 3) Observational Studies. If conducted and interpreted correctly, these types of research can help to inform health care decision-making. If, however, such studies are conducted or interpreted incorrectly, the comparative answers from these studies may be inaccurate, or worse, misleading. This white paper will review the three CER study types, discuss their proper methodologic approaches, and provide tools for CER consumers to use when they review them. Although additional types of biomedical research are very useful in understanding disease and developing new drugs, the focus of this paper is solely on the most common types of research used for CER.

Comparative Effectiveness Research

CER is not a new tool, and it has already produced many well publicized results regarding commonly used therapies. For example, the ALLHAT clinical trial, a government-funded $130 million study, compared a diuretic with two newer (at the time) blood pressure-lowering drugs to determine their effects on cardiovascular (CV) disease.[2] The study found that for single-drug initial therapy, thiazide-type diuretics on average delivered superior blood-pressure control, reduced CV events, and had better patient tolerance compared with the comparators.[3]

In another example, CATIE, a $42.6 million study sponsored by the National Institute of Mental Health, compared an older antipsychotic to several second-generation drugs.[4] The CATIE trial found that the first-generation antipsychotic delivered overall effectiveness similar to the other drugs. However, a very large number of patients switched to an alternative drug during the study, suggesting that for an individual patient the overall results may not apply.[5]

CER also examines non-drug interventions: a recent (2009) study compared percutaneous coronary intervention (PCI) to coronary-artery bypass grafting (CABG) in the treatment of severe coronary artery disease. The study found that while strokes occurred more frequently following CABG, rates of major adverse cardiac or cerebrovascular events were significantly higher in PCI. The authors concluded that CABG should remain the standard of care in indicated patients.[6]

Despite the large costs, complexity, and size of these important comparative effectiveness clinical trials, some commentators have: a) raised issues about their design, b) voiced concerns that the results have been extrapolated beyond what was intended, and c) questioned their current relevance because of newly available drugs and a broader understanding of the underlying diseases and their treatment. To impact care, CER needs to anticipate and subsequently address these critiques.[4,7,8]

More comparative effectiveness data will soon become available as the American Recovery and Reinvestment Act of 2009 allocated $1.1 billion toward this effort. Understanding the design and interpretation of CER will allow the reader to more fully assess the applicability, rigor, and conclusions of forthcoming reports.

How to Approach CER

To conduct a comparative study, the researcher first matches the design to the questions needing answers and the constraints placed upon him (e.g., financial, required timeline, ethical and practical considerations). For example, what study design will best investigate my outcomes of interest? What study duration can assess the key efficacy and safety outcomes? How much funding is available for research? Randomized controlled trials, meta-analyses, and observational studies will form the bulk of the new CER. An understanding of each study's inherent strengths and weaknesses is essential for an informed CER consumer.

RCTs are experiments wherein subjects in a population are randomly allocated into study and control groups to receive or not to receive an experimental intervention such as a new therapeutic biopharmaceutical or medical device. Although some RCTs include large numbers of patients to address broad population questions about competing therapies, studies of this type often apply the intervention to a carefully selected population. This more narrowly defined patient group focuses attention on the people at high risk for a disease who are most likely to benefit from its treatment while reducing the impact of external factors (such as comorbidities) that could also influence the outcome of interest. The necessary controls and careful patient selection for these investigations means that their findings may not apply to a broader population. Even so, the strengths associated with RCT methodology, particularly in its lowered risk of bias and confounding, have allowed this type of study design to gain sufficient credibility that many now consider it to be the "gold

standard" of clinical research. RCTs are required by the Food and Drug Administration (FDA) for drug approval and are designed to achieve the most robust scientific data about the safety and efficacy of an intervention.

Meta-analyses synthesize data across a series of similar trials, generally RCTs. This study design requires careful selection of which trials to include and which statistical methods to employ. Publication bias, which occurs when positive studies are more abundantly available on a topic due to unpublished negative studies, can significantly impair the validity of this study type. However, a balanced meta-analysis can produce a summary estimate of the medical literature that underpins benchmarks of clinical decision-making.

Observational studies follow a cohort of participants over a period of time to examine the potential associations between patients' exposure to treatment and health outcomes. These studies can be performed prospectively, observing patients in real time, or they can be retrospective analyses of existing databases. Given the abundance of billing or administrative databases, comparisons of therapies and their outcomes will frequently use these databases, as they embody the experience of patients in real-world settings. While results from observational studies tend to generate hypotheses that require the more rigorous RCT for confirmation, long-term observational studies have provided substantial medical knowledge (e.g., demonstrating the relationship between elevated blood cholesterol and the development of heart disease).

Several themes apply to all three study designs. These themes and a framework for their assessment will be explored in the case studies to follow.

STEP 1: Consider for whom the findings are applicable.

Typically, the assessment of new research begins with a careful consideration of the quality of the study, or its internal validity. Alternatively, first assessing its external validity by relating the study to your own needs may be the most efficient course of action when interpreting findings of CER.

If the study used a population, intervention, or outcomes that are not relevant to your own situation, there is little need to continue further. For example, a study investigating the effect of the H1N1 vaccine in a school-age population may not help an internist seeking evidence to inform the care of older individuals. More broadly, the reader of a new CER report should first determine whether the available study can be generalized to the reader's environment of interest. To help in assessing applicability, consider the following questions (see Table 1 for a more complete list):

- Does the study population resemble your population of interest (e.g., age, gender, socioeconomic status, disease profile, illness severity)?

- Does the type of clinical environment resemble your own (e.g., primary care, inner city)?
- Is the study conducted in a "real-world" setting or a highly controlled homogeneous population?
- Are the outcomes being measured appropriate to your needs in making health care decisions (e.g., short vs. long term, quality of life, physical function, frequency of hospitalization)?

Consider the study in light of these questions. Keep in mind that there is no cutoff score (e.g., if there are > four "yes" boxes checked, then the study has relevance). A single issue may invalidate the study from consideration (e.g., the study was too short to assess my outcomes of interest). Alternatively, this study even with marked limitations may provide adequate evidence to support a clinical choice. These evaluation questions should guide your review, not determine the final assessment.

STEP 2: Consider whether there are any aspects of the study design that might greatly affect the results.

If the study has relevance (passes Step 1), then closely examine the study's validity. To do so, first consider whether the researchers selected a study design that matches the question being asked. For example, why did the authors conduct another RCT when a more real-world environment would better answer the research question? Will the population in an observational study likely have so much patient heterogeneity that confounding cannot be avoided? As an example, in routine care, physicians may prescribe a new antibiotic only for their sickest patients. In that circumstance, the antibiotic's real benefit may be masked since those sicker patients would ordinarily do worse than the healthier patients prescribed the alternative treatment. Finally, are meta-analyses being conducted just because studies are available for synthesis or do the available studies truly address the question asked?

After assessing the choice of study design, examine whether that design was carried out with adequate rigor. The issues involved in making that assessment will differ depending upon the methodology used in the CER. For example, RCTs should be assessed for methodological issues such as blinding and randomization techniques and the statistical tests employed to account for population changes in the study arms. Given the scope of this question, this white paper cannot review all of the threats to validity in an RCT. Many clinical epidemiology texts can provide that detail. In brief form, Table 2 contains selected questions to consider. Although not a guarantee, articles from important journals undergo rigorous peer

review. That review should identify the most critical threats to internal validity. Chapter 2 will explore this topic more fully.

In a meta-analysis, the results strongly depend upon the studies chosen for inclusion and whether it was appropriate to combine results across them. What were the preset study inclusion/exclusion criteria? Were the studies sufficiently similar in design (e.g., intervention, dose, comparators, duration, and outcomes) that a quantitative synthesis would be valid? Finally, the statistical methods can also greatly influence the results. Although a non-statistician may have difficulty critiquing those choices, a well-written document should discuss why the authors selected the methods they used and whether the results were sensitive to those choices. A non-statistician can and should review that discussion. See Table 3 and Chapter 3 for additional considerations.

Observational studies are often the easiest to perform (especially if the researchers analyze an existing database). However, that ease may lead to a large volume of analyses with inherent biases or statistical aberrations. By design, these studies are not controlled experiments, they do not randomize patients, and findings in these studies may reflect peculiarities of the database or unmeasured characteristics of the patients. In reviewing these studies, it is important to understand whether the authors had a pre-determined hypothesis, how they addressed potential biases, and how they explored limitations in their data (see Table 4). Chapter 4 will present an observational case study and address these and other issues.

STEP 3: Consider whether the findings may change with new research.

After completing these two steps (i.e., relevance and rigor), the reader should consider whether the apparently definitive conclusion may change over time. The inherent nature of research is cumulative, wherein new findings constantly emerge to build on older results. As such, all findings should be treated with that knowledge and with the expectation that further results may appear that expand upon, clarify, or reject these findings. Keeping the patient in mind, perhaps adopt the mindset: "How sure are we, really?" To answer this question, consider whether the area of research changes rapidly. If so, it raises the likelihood that advances in scientific knowledge will alter the current findings. For example, an increased understanding of the biology and genetics of cancer will probably influence the true benefits and harms in an ever-evolving target population.

Consider also the influence of the CER upon clinical policy. If the findings will profoundly affect access to a therapy or promulgated standards of care, greater confidence in the stability of the findings would be important. If the area of medicine has relative stability and the impact of the research has a more modest

purview, the reader may accept less certainty. Table 5 provides a broader check list of considerations on this topic and Chapter 5 provides additional information.

The Case Studies

The next three chapters provide examples of comparative effectiveness research. Each chapter addresses a separate research design: Chapter 2, randomized controlled trials; Chapter 3, meta-analysis; and Chapter 4, observational studies. Each begins with a brief summary of the study design, followed by a presentation of a clinical situation and a particular therapy. Using high profile studies, each chapter raises issues to consider and how those issues may influence the results. The final chapter reconsiders the three review steps in light of the lessons learned.

Having finished this monograph, the reader will have the tools to review a new CER report. In addition, the case studies will provide lessons learned and tips for each study design type. Combining the tools and tips will enable the reader to determine whether a new CER report can or should influence the reader's decision-making.

TABLE 1: ASSESSING EXTERNAL VALIDITY		
1. Does the study population reflect your population of interest?	___yes	___no
2. Does the type of clinical environment reflect your own?	___yes	___no
3. Is the study conducted in a "real-world" setting?	___yes	___no
4. Are the outcomes being measured appropriate to your needs?	___yes	___no
5. Is the evidence clinically significant, and the effect size large enough to be relevant?	___yes	___no
6. Does the duration of the study allow you to draw any long-term conclusions?	___yes	___no
7. Have appropriate subgroups been considered?	___yes	___no

TABLE 2: RANDOMIZED CONTROLLED TRIALS		
1. Is there a clear hypothesis?	___yes	___no
2. Were the inclusion criteria logical?	___yes	___no
3. Was the comparison group appropriate?	___yes	___no
4. Were study methods clearly described and valid?	___yes	___no
5. Are all relevant outcomes reported?	___yes	___no
6. Is there a strong exposure/outcome link?	___yes	___no
7. Are study limitations discussed?	___yes	___no

TABLE 3: META-ANALYSIS		
1. Is there a clear hypothesis?	___yes	___no
2. Were the outcomes clearly defined and clinically relevant?	___yes	___no
3. Were the inclusion criteria logical?	___yes	___no
4. Were all relevant studies included?	___yes	___no
5. Do the authors assess the potential for publication bias?	___yes	___no
6. Was the search for relevant studies thorough?	___yes	___no
7. Were multiple searchable databases employed?	___yes	___no
8. Were the included studies of high methodologic quality?	___yes	___no
9. Did the authors test whether their studies were similar (homogeneous) or different (heterogeneous)?	___yes	___no
10. Were the combined studies sufficiently similar?	___yes	___no
11. Did the authors adequately describe and defend the statistical methods they used for analysis?	___yes	___no

TABLE 4: OBSERVATIONAL STUDIES		
1. Does the population seem properly defined for the stated outcomes of interest?	___yes	___no
2. Are the baseline characteristics similar for the groups being studied?	___yes	___no
3. Were study methods clearly described and valid?	___yes	___no
4. Are all relevant outcomes reported?	___yes	___no
5. Is there a strong exposure/outcome link?	___yes	___no
6. Did the authors adequately describe and defend the statistical methods they used for analysis?	___yes	___no
7. Have potentials for bias and confounding been discussed and accounted for?	___yes	___no

TABLE 5: INNOVATION AND CHANGE*		
1. Is this research in an area that is changing rapidly (such as cancer research)?	___yes	___no
2. Are advances in scientific knowledge and understanding likely to influence how we view the effectiveness of this intervention?	___yes	___no
3. Will these results change clinical guidelines or impact policy?	___yes	___no
4. Is future work in this arena likely to take place from existing databases?	___yes	___no
5. Could subgroup analysis clarify these results?	___yes	___no
** In contrast to the prior tables, a "yes" indicates that the reader should view the study results with caution.*		

A note on the reporting of study results seen in this monograph:

The **P value** assesses the strength of the study results, or more specifically whether a difference that is observed is likely to be statistically valid (versus occurring by chance alone). Researchers often set $P<0.05$ as their cutoff for statistical significance, which means that there is less than a 5% chance that the result occurred by chance.

Relative Risk (RR) is commonly used to compare the likelihood of an event occurring between two groups. If the risk of a heart attack is 8% in one group and 10% in the comparison group, the RR is 0.8 (or 8%/10%). An **Odds Ratio** is a similar concept that is used in specific study designs. When the events are uncommon, the relative risk and the odds ratios are essentially the same. **Hazard Ratios**, which will be seen in the cetuximab case study, are an estimate of RR.

Confidence intervals (CI) provide a range of values around the "average" result. Commonly, studies report a **95% CI**. By providing such a range, the study investigator is suggesting that 95% of the time, the study result will fall within that range.

CHAPTER 2: LESSONS LEARNED ABOUT RANDOMIZED CONTROLLED TRIALS: THE CETUXIMAB CASE STUDY

Background

What is a randomized controlled trial?

Randomized controlled trials answer the question: "Can an intervention work under certain controlled conditions?" To do so, an RCT employs a carefully planned experimental framework to compare an intervention with a control, investigating the effect of each treatment option on a defined outcome. After meeting strict inclusion qualifications, each participant is randomly assigned to one of the treatment groups. Ideally, the study is "double-blinded," meaning that neither the participants nor the researchers know who is receiving the active intervention and who is receiving the control treatment. RCTs are "single" blinded when only the participants are unaware of their allocation. Appropriate randomization ensures that the groups are similar, so that observed effects are unlikely to be due to differences between those groups. In this study type, the investigator narrowly defines the patients, the interventions, and the outcomes to determine the benefits and harms of the intervention in an ideal setting. RCTs may compare an intervention to placebo or to another active therapy. Although CER focuses on comparisons of alternative therapies, the placebo-controlled RCT provides important information about the magnitude of a drug's benefits and frequency of adverse reactions. These positive and negative outcomes may be over- or under-estimated in head-to-head randomized, controlled, active comparator studies. Moreover, placebo-controlled trials are also a key step in the route to gaining FDA approval for new therapies, or new indications for existing therapies.

Of all clinical research study designs, RCTs provide the strongest internal validity: (i.e., this drug, taken at this dose, produces this outcome in this group of people). The randomization and blinding techniques imposed at the start of a study greatly reduce the likelihood that bias or confounding influences the results. Confounding occurs when external factors are more likely in one study arm and have an independent impact on the measured outcomes (e.g., the intervention group has sicker patients than the control group). Because they reduce the impact of confounding, RCTs are considered the gold standard of clinical research, and many policymakers hope that there will be an increasing number of RCTs as CER continues to grow.

However, RCTs have various potential limitations. First, the very characteristics that support an RCT's strong internal validity can limit the study design's external validity, or generalizability to the broader world of care beyond the study's controlled environment. RCTs generally enroll a narrow set of patients without relevant comorbidities and rarely include children or the elderly in their study populations. Second, most RCTs are modest in size due to their expense and challenging logistics. This modest size limits RCTs' ability to assess

uncommon events. Third, many RCTs compare active treatments with placebo, rather than head-to-head with another active treatment, or choose to use comparators that may not be "right" or "best" for the analysis. Finally, the strict pre-set design of RCTs is occasionally abandoned by participants who may find themselves unhappy with their randomization or the constraints of the study; participants leaving a study (known as loss to follow-up) or requiring a treatment that might necessitate their transfer from the placebo to intervention arm may diminish the statistical power of the study.

Even with its disadvantages, the RCT (along with meta-analyses of multiple RCTs) provides the highest level of evidence.[9] The following case study will illustrate several characteristics of this study design that will aid in evaluating newly published work.

The Clinical Situation

Colorectal cancer is the third most commonly diagnosed cancer and the third leading cause of cancer death in both men and women in the U.S.[10] The 5-year survival for patients diagnosed in the metastatic stage (mCRC) is 11.3%, compared to 90.8% in patients diagnosed when the tumor is still localized.[11] Treatment for mCRC rarely cures the disease; its aim is palliative, intending to decrease tumor-related symptoms.[12] Although non-curative, various chemotherapy regimens have improved the median survival of mCRC patients from 12 months to approximately 18-21 months.[13] Basic research has identified the molecular or genetic basis of cancer and how it affects disease progression. This understanding of cancer pharmacogenomics has highlighted the opportunity to test patients (e.g., employing biomarkers) and the potential ability to target therapies to those most likely to benefit.

Randomized Controlled Trials

What was done and what was found?

Cetuximab, an anti-epidermal growth factor receptor (EGFR) agent, has recently been added to the therapeutic armamentarium. Two important RCTs examined its impact in patients with mCRC. In the first (2004), 56 centers in 11 European countries investigated the outcomes associated with cetuximab therapy in 329 mCRC patients who experienced disease progression either on irinotecan therapy or within 3 months thereafter.[13] The authors found that the group on a combination of irinotecan and cetuximab had a significantly higher rate of overall response to treatment (primary endpoint) than the group on cetuximab alone: 22.9% (95% CI, 17.5-29.1%) vs. 10.8% (95% CI, 5.7-18.1%) (P=0.007), respectively. Similarly, the median time to progression was significantly longer in the combination therapy group (4.1 vs. 1.5 months, P<0.001). Since these patients had already progressed on irinotecan prior to the study, any response was viewed as positive. Safety between the two treatment arms was similar: approximately 80% of patients in each arm experienced a rash. Grade 3 or 4 (the more severe) toxic effects

on the skin were slightly more frequent in the combination-therapy group compared to cetuximab monotherapy, observed in 9.4% and 5.2% of participants, respectively. Other side effects such as diarrhea and neutropenia observed in the combination-therapy arm were considered to be in the range expected for irinotecan alone. Data from this study demonstrated the efficacy and safety of cetuximab and were instrumental in the FDA's 2004 approval.[14]

A second RCT (2007), of 572 patients, also showed the efficacy of cetuximab in the treatment of mCRC. The study was a randomized, non-blinded, controlled trial that examined cetuximab monotherapy plus best supportive care compared to best supportive care alone in patients who had received and failed prior chemotherapy regimens.[15] This study found that median overall survival (the primary endpoint) was significantly higher in patients receiving cetuximab plus best supportive care compared to best supportive care alone (6.1 vs. 4.6 months, respectively) (hazard ratio for death=0.77; 95% CI: 0.64-0.92, P=0.005). This RCT described a greater incidence of adverse events in the cetuximab plus best supportive care group compared to best supportive care alone including (most significantly) rash, as well as edema, fatigue, nausea and vomiting.

Was this the right answer?

The RCTs discussed thus far had fairly broad enrollment criteria and the cetuximab benefits were modest. Emerging scientific theories raised the possibility that genetically defined population subsets might experience a greater-than-average treatment benefit. One such area of inquiry entailed examining "biomarkers," or genetic indicators of a patient's greater response to therapy. Even as the above RCTs were being conducted, data emerged showing the importance of the KRAS gene.

Emerging Data

Based on the emerging biochemical evidence that the EGFR treatment mechanism was even more finely detailed than previously understood, the study authors of the 2007 RCT undertook a retrospective subgroup analysis using tumor tissue samples preserved from their initial study.[16] Following laboratory analysis, all viable tissue samples were classified as having a wild-type (non-mutated) or a mutated KRAS gene. Instead of the previous two study arms (cetuximab plus best supportive care vs. best supportive care alone), there were four for this new analysis: each of the two original study arms was further divided by wild-type vs. mutated KRAS status.

Laboratory evaluation determined that 40.9% and 42.3% of all patients in the RCT had a KRAS mutation in the cetuximab plus best supportive care group compared to the best supportive care group alone, respectively. The efficacy of cetuximab was found to be significantly correlated with KRAS status: in patients with wild-type (non-mutated)

KRAS genes, cetuximab plus best supportive care compared to best supportive care alone improved overall survival (median 9.5 vs. 4.8 months, respectively; hazard ratio for death=0.55; 95% CI, 0.41-0.74, P<0.001), and progression-free survival (median 3.7 vs. 1.9 months, respectively; hazard ratio for progression or death=0.40; 95% CI, 0.30-0.54, P<0.001). Meanwhile, in patients with mutated KRAS tumors, the authors found no significant difference in outcome between cetuximab plus best supportive care vs. best supportive care alone.

Where to go from here?

Based on these and similar results from other studies, the FDA narrowed its product labeling in July 2009 to indicate that cetuximab is not recommended for mCRC patients with mutated KRAS tumors.[17] This distinction reduces the relevant population by approximately 40%.[17,18] Similarly, the American Society of Clinical Oncology released a provisional clinical recommendation that all mCRC patients have their tumors tested for KRAS status before receiving anti-EGFR therapy.[19]

The benefits of targeted treatment are many. Patients who previously underwent cetuximab therapy without knowing their genetic predisposition would no longer have to be exposed to the drug's toxic effects if unnecessary, as the efficacy of cetuximab is markedly higher in the genetically defined appropriate patients. In a less-uncertain environment, clinicians can be more confident in advocating a course of action in their care of patients. And finally, knowledge that targeted therapy is possible suggests the potential for further innovation in treatment options. In fact, research continues to demonstrate options for targeted cetuximab treatment of mCRC at an even finer scale than seen with KRAS[20-22]; and similar genetic targeting is being investigated, and advocated, in other cancer types.[23]

Lessons Learned From This Case Study

Although RCTs are generally viewed as the gold standard, results of one or even a series of trials may not accurately reflect the benefits experienced by an individual patient. This case study showed that cetuximab initially appeared to have rather modest clinical benefits. However, newer information became available and subsequent genetic subgroup assessments led to very different conclusions. Clinicians should be aware that the current knowledge is likely to evolve and any decisions about patient care should be carefully considered with that sense of uncertainty in mind.

As in this case study, subgroup analyses (e.g., genetic subtypes) need a theoretical rationale. Ideally, the analyses should be determined at the time of original RCT design and should not just occur as explorations of the subsequent data. Done improperly, post hoc analyses may lead to incorrect patient care conclusions.

Randomized Controlled Trials: Tips for the CER Consumer

- RCTs can determine whether an intervention can provide benefit in a very controlled environment.
- The controlled nature of an RCT may limit its generalizability to a broader population.
- No results are permanent; advances in scientific knowledge and understanding can influence how we view the effectiveness (or safety) of a therapeutic intervention.
- Targeted therapy illuminated by carefully thought out subgroup analyses can improve the efficacious and safe use of an intervention.

CHAPTER 3: LESSONS LEARNED ABOUT META-ANALYSIS: THE ROSIGLITAZONE CASE STUDY

Background

What is a meta-analysis?

Not infrequently, results for the same intervention differ across clinical trials and it may not be clear whether one therapy provides more benefit than another. As CER increases and more studies are conducted, clinicians and policymakers are more likely to encounter this scenario. In a systematic review, a researcher identifies similar studies and displays their results in a table, enabling qualitative comparisons across the studies. With a meta-analysis, the data from included studies are statistically combined into a single "result." Merging the data from a number of studies increases the effective sample size of the investigation, providing a statistically stronger conclusion about the body of research. By so doing, investigators may detect low frequency events and demonstrate more subtle distinctions between therapeutic alternatives.

When studies have been properly identified and combined, the meta-analysis produces a summary estimate of the findings and a confidence interval that can serve as a benchmark in medical opinion and practice. However, when done incorrectly, the quantitative and statistical analysis can create impressive "numbers" but biased results. The following are important criteria for properly conducted meta-analyses:

1. Carefully defining unbiased inclusion or exclusion criteria for study selection
2. Including only those studies that have similar design elements, such as patient population, drug regimen, outcomes being assessed, and timeframe
3. Applying correct statistical methods to combine and analyze the data

Reporting this information is essential for the reader to determine whether the data were suitable to combine, and if the meta-analysis draws unbiased conclusions.

Meta-analyses of randomized clinical trials are considered to be the highest level of medical evidence as they are based upon a synthesis of rigorously controlled trials that systematically reduce bias and confounding. This technique is useful in summarizing available evidence and will likely become more common in the era of publicly funded comparative effectiveness research. The following case study will examine several key principles that will be useful as the reader encounters these publications.

The Clinical Situation

Heart disease is the leading cause of mortality in the United States, resulting in approximately 20% of all deaths.[24] Diabetics are particularly susceptible to heart disease,

with more than 65% of deaths attributable to it.[25] The nonfatal complications of diabetes are wide-ranging and include kidney failure, nerve damage, amputation, stroke and blindness, among other outcomes.[26] In 2007, the total estimated cost of diabetes in the United States was $174 billion; $116 billion was derived from direct medical expenditures and the rest from the indirect cost of lost productivity due to the disease.[27] With such serious health effects and heavy direct and indirect costs tied to diabetes, proper disease management is critical. Historically, diabetes treatment has focused on strict blood sugar control, assuming that this goal not only targets diabetes but also reduces other serious comorbidities of the disease.[28]

Anti-diabetic agents have long been associated with key questions as to their benefits/risks in the treatment of diabetes. The sulfonylurea tolbutamide, a first generation anti-diabetic drug, was found in a landmark study in the 1970s to significantly increase the CV mortality rate compared to patients not on this agent.[29] Further analysis by external parties concluded that the methods employed in this trial were significantly flawed (e.g., use of an "arbitrary" definition of diabetes status, heterogeneous baseline characteristics of the populations studied, and incorrect statistical methods).[30,31] Since these early studies, CV concerns continue to be an issue with selected oral hypoglycemic agents that have subsequently entered the marketplace.

The thiazolidinedione (TZD) drug class, approved in the late 1990s, initially seemed to be a solution to the problems associated with the older generation of sulfonylureas. Rosiglitazone, a member of the TZD class, was approved by the FDA in 1999 and was widely prescribed for the treatment of type 2 diabetes. A number of RCTs supported the benefit of rosiglitazone as an important new oral antidiabetic agent.[32-37] However, safety concerns developed as the FDA received reports of adverse cardiac events potentially associated with rosiglitazone.[38,39] It was in this setting that a meta-analysis by Nissen and Wolski was published in the *New England Journal of Medicine* in June 2007.[40]

Meta-Analysis

What was done?

Nissen and Wolski conducted a meta-analysis examining the impact of rosiglitazone on cardiac events and mortality compared to alternative therapeutic approaches. The authors began with a broad search to locate potential studies for review. They screened published phase II, III, and IV trials; the FDA website; and the drug manufacturer's clinical-trial registry for applicable data relating to rosiglitazone use. When the initial search was complete, the studies were further categorized by pre-stated inclusion criteria. The meta-analysis inclusion criteria were simple: studies had to include rosiglitazone and a randomized comparator group treated with either another drug or placebo, study arms had to show similar length of treatment, and all groups had to have received more than 24 weeks of exposure to

the study drugs. Perhaps most important, studies had to contain outcome data of interest including the rate of myocardial infarction (MI) or death from all CV causes. Out of 116 studies surveyed by the authors, 42 met their inclusion criteria and were included in the meta-analysis. Of the studies they included, 23 had durations of 26 weeks or less, and only five studies followed patients for more than a year.

Until this point, the study's authors were following a path similar to that of any reviewer interested in CV outcomes, examining the results of these 42 studies and comparing them qualitatively. Quantitatively combining the data, however, required the authors to make choices about the studies they could merge and the statistical methods they should apply for analysis. Those decisions greatly influenced the results that were reported.

What was found?

When the studies were combined, the meta-analysis contained data from 15,565 patients in the rosiglitazone group and 12,282 patients as comparators. To analyze their data, the authors chose one particular statistical method (the Peto method, a fixed-effect statistical approach), which calculates the odds of events occurring where the outcomes of interest are rare and small in number.[41]

In comparing rosiglitazone with a "control" group that included other drugs or placebo, the authors reported odds ratios of 1.43 (95% CI, 1.03-1.98; P=0.03) and 1.64 (95% CI, 0.98-2.74; P=0.06) for MI and death from CV causes, respectively. In other words, the odds of an MI or death from a CV cause are higher for rosiglitazone patients than for patients on other therapies or placebo. The authors reported that rosiglitazone was significantly associated with an increase in the risk of MI and had borderline significance in increasing the risk of death from all CV causes.[40] These findings appeared online on the same day that the FDA issued a safety alert regarding rosiglitazone.[42] Discussion of the meta-analysis was immediately featured prominently in the news media.[43,44] By December, prescription claims for the drug at retail pharmacies had fallen by more than 50%.[43]

As diabetic patients and their clinicians reacted to the news, a methodologic debate also ensued. This discussion included statistical issues pertaining to the conduct of the analysis, its implications for clinical care, and finally the FDA and drug manufacturer's roles in overseeing and regulating rosiglitazone.[38,45-50] The debate, as well as concern among patients with diabetes regarding treatment, continues in the medical community today.

Was this the right answer?

Should the studies have been combined?

Commentators faulted the authors for including several studies that were not originally intended to investigate diabetes,[51] and for combining both placebo and drug therapy data into one comparator arm.[47] Some critics noted that despite the stated inclusion criteria,

some data were derived from studies where the rosiglitazone arm was allowed a longer follow-up than the comparator arm. By failing to account for this longer follow-up period, commentators felt that the authors may have overestimated the effect of rosiglitazone on CV outcomes.[52] Many reviewers were concerned that this meta-analysis excluded trials in which no patients suffered an MI or died from CV causes – the outcomes of greatest interest. Some reviewers also noted that the exclusion of zero-event trials from the pooled dataset not only gave an incomplete picture of the impact of rosiglitazone but could have increased the odds ratio estimate.[53] In general, the pooled dataset was criticized by many for being a faulty microcosm of the information available regarding rosiglitazone.

It is essential that a meta-analysis be based on similarity in the data sources. If studies differ in important areas such as the patient populations, interventions, or outcomes, combining their data may not be suitable.[54] The researchers accepted studies and populations that were clinically heterogeneous, yet pooled them as if they were not. The authors noted that the results were combined from a number of trials that were not initially intended to investigate CV outcomes. Furthermore, the available data did not allow for time-to-event analysis, an essential tool in comparing the impact of alternative treatment options. Reviewers considered the data to be insufficiently homogeneous, and the line of cause and effect to be murkier than the authors described.

Were the statistical methods optimal?

The statistical methods for this meta-analysis also came under significant criticism. The authors' use of the Peto method was criticized for being an incorrect choice because data were pooled from both small and very large studies, resulting in a potential overestimation of treatment effect.[55] Others noted that the Peto method should not have been used, as a number of the underlying studies did not have patients assigned equally to rosiglitazone and comparator groups.[56] Finally, critics suggested that the heterogeneity of the included studies required an altogether different set of analytic techniques.

Demonstrating the sensitivity of the authors' initial analysis to the inclusion criteria and statistical tests used, a number of researchers reworked the data from this study. One researcher used the same studies but analyzed the data with a more commonly used statistical method (Mantel-Haenszel), and found no significant increase in the relative risk or common odds ratio with MI or CV death.[56] When the pool of studies was expanded to include those originally eliminated because they had zero CV events, the odds ratios for MI and death from CV causes dropped from 1.43 to 1.26 (95% CI, 0.93-1.72) and from 1.64 to 1.14 (95% CI, 0.74-1.74), respectively. Neither of the recalculated odd ratios were significant for MI or CV death.[57] Finally, several newer long-term studies have been published since the Nissen meta-analysis. Incorporating their results with the meta-analysis data showed that rosiglitazone is associated with an increased risk of MI but not of CV

death.[55] Thus, the findings from these meta-analyses varied with the methods employed, the studies included, and the addition of later trials.

Emerging Data

The controversy surrounding the rosiglitazone meta-analysis authored by Nissen and Wolski forced an unplanned interim analysis of a long-term, randomized trial investigating the CV effects of rosiglitazone among patients with type 2 diabetes. The authors of the RECORD trial noted that even though the follow-up at 3.75 years was shorter than expected, rosiglitazone, when added to standard glucose-lowering therapy, was found to be associated with an increase in the risk of heart failure but was not associated with any increase in death from CV or other causes. Data at the time were found to be insufficient to determine the effect of rosiglitazone on an increase in the risk of MI.[58] The final report of that trial, published in June 2009, confirmed the elevated risk of heart failure in people with type 2 diabetes treated with rosiglitazone in addition to glucose-lowering drugs, but continued to show inconclusive results about the effect of the drug therapy on the risk of MI. Further, the RECORD trial clarified that rosiglitazone does not result in an increased risk of CV morbidity or mortality compared to standard glucose-lowering drugs.[59] Other trials conducted since the publishing of the meta-analysis have corroborated these results, casting further doubt on the findings of the meta-analysis published by Nissen and Wolski.[55,60]

Where to go from here?

Some sources suggest that the original Nissen meta-analysis delivered more harm than benefit, and that a well-recognized medical journal may have erred in its process of peer review.[51] Despite this criticism, it is important to note that subsequent publications support the risk of adverse CV events associated with rosiglitazone, although rosiglitazone use does not appear to increase deaths.[55,59,60] Their results and emerging data point to the need for further rigorous research to clarify the benefits and risks of rosiglitazone on a variety of outcomes, and the importance of directing the drug to the population that will maximally benefit from its use.

Lessons Learned From This Case Study

Results from initial randomized trials that seem definitive at one time may not be conclusive, as further trials may emerge to clarify, redirect, or negate previously accepted results. A meta-analysis of those trials can lead to varying results based upon the timing of the analysis and the choices made in its performance.

Meta-Analysis: Tips for the CER Consumer

- The results of a meta-analysis are highly dependent on the studies included (and excluded). Are these criteria properly defined and relevant to the purposes of the meta-analysis? Were the combined studies sufficiently similar? Can results from this cohort be generalized to other populations of interest?
- The statistical methodology can impact study results. Have there been reviews critiquing the methods used in the meta-analysis?
- A variety of statistical tests should be considered, and perhaps reported, in the analysis of results. Do the authors mention their rationale in choosing a statistical method? Do they show the stability of their results across a spectrum of analytical methods?
- Nothing is permanent. Emerging data may change the playing field, and meta-analysis results are only as good as the data and statistics from which they are derived.

CHAPTER 4: LESSONS LEARNED ABOUT OBSERVATIONAL STUDIES: THE NURSES' HEALTH STUDY

Background

What is an observational study?

An observational study is a very common type of research design in which the effects of a treatment or condition are studied without formally randomizing patients in an experimental design. These studies can be done prospectively, wherein data are collected about a group of patients going forward in time; or retrospectively, in which the researcher looks into the past, mining existing databases for data that have already been collected. These latter studies are frequently performed by using an electronic database that contains, for example, administrative, "billing," or claims data. Less commonly, observational research uses electronic health records, which have greater clinical information that more closely resembles the data collected in an RCT. Observational studies often take place in "real-world" environments, which allow researchers to collect data for a wide array of outcomes. Patients are not randomized in these studies, but the findings can be used to generate hypotheses for investigation in a more constrained experimental setting. Perhaps the best known observational study is the "Framingham Study," which collected demographic and health data for a group of individuals over many years (and continues to do so) and has provided an understanding of the key risk factors for heart disease and stroke.[61]

Observational studies present many advantages to the comparative effectiveness researcher. The study design can provide a unique glimpse of the use of a health care intervention in the "real world," an essential step in gauging the gap between efficacy (can a treatment work in a controlled setting?) and effectiveness (does the treatment work in a real-life situation?). Furthermore, observational studies can be conducted at low cost, particularly if they involve the secondary analysis of existing data sources. CER often uses administrative databases, which are based upon the billing data submitted by providers during routine care. These databases typically have limited clinical information, may have errors in them, and generally do not undergo auditing.

The uncontrolled nature of observational studies allows them to be subject to bias and confounding. For example, doctors may prescribe a new medication only for the sickest patients. Comparing these outcomes (without careful statistical adjustment) with those from less ill patients receiving alternative treatment may lead to misleading results. Observational studies can identify important associations but cannot prove cause and effect. These studies can generate hypotheses that may require RCTs for fuller demonstration of those relationships. Secondary analysis can also be problematic if researchers overwork datasets by doing multiple exploratory analyses (e.g., data-dredging): the more we look, the more we find, even if those findings are merely statistical aberrations. Unfortunately,

the growing need for CER and the wide availability of administrative databases may lead to selection of research of poor quality with inaccurate findings.

In comparative effectiveness research, observational studies are typically considered to be less conclusive than RCTs and meta-analyses. Nonetheless, they can be useful, especially because they examine typical care. Due to lower cost and improvements in health information, observational studies will become increasingly common. The reader will need to critically assess whether the described results are helpful or biased based upon how the study was performed. This case will illustrate several characteristics of the types of studies that will assist in evaluating newly published work.

The Clinical Situation

Cardiovascular diseases (CVD) are the leading cause of death in women older than the age of 50.[62] Epidemiologic evidence suggests that estrogen is a key mediator in the development of CVD. Estrogen is an ovarian hormone whose production decreases as women approach menopause. The steep increase in CVD in women at menopause and older and in women who have had hysterectomies further supports a relationship between estrogen and CVD.[63,64] Building on this evidence of biologic plausibility, epidemiological and observational studies suggested that estrogen replacement therapy (a form of hormone replacement therapy, or HRT) had positive effects on the risk of CVD in postmenopausal women, (albeit with some negative effects in its potential to increase the risk for breast cancer and stroke).[65] Based on these findings, in the 1980s and 1990s HRT was routinely employed to treat menopausal symptoms and serve as prophylaxis against CVD.[66]

Observational Study

What was done?

The Nurses' Health Study (NHS) began collecting data in 1976. In the study, researchers intended to examine a broad range of health effects in women over a long period of time, and a key goal was to clarify the role of HRT in heart disease. The cohort (i.e., the group being followed) included married registered nurses aged 30-55 in 1976 who lived in the 11 most populous states. To collect data, the researchers mailed the study participants a survey every 2 years that asked questions about topics such as smoking, hormone use, menopausal status, and less frequently, diet. Data were collected for key end points that included MI, coronary-artery bypass grafting or angioplasty, stroke, total CVD mortality, and deaths from all causes.

What was found?

At a 10-year follow-up point, the NHS had a study pool of 48,470 women. The researchers found that estrogen use (alone, without progestin) in postmenopausal women was associated with a reduction in the incidence of CVD as well as in CVD mortality compared

to non-users.[67] Later, estrogen-progestin combination therapy was shown to be even more cardioprotective than estrogen monotherapy,[68] and lower doses of estrogen replacement therapy were found to deliver equal cardioprotection and lower the risk for adverse events.[69] NHS researchers were alert to the potential for bias in observational studies. Adjustment for risk factors such as age (a typical practice to eliminate confounding) did not change the reported findings.[67]

Was this the right answer?

The NHS was not unique in reporting the benefits associated with HRT; other observational studies corroborated the NHS findings. A (retrospective) secondary data analysis of the UK primary care electronic medical record database, for example, also showed the protective effect associated with HRT use.[70] Nonetheless, researchers were aware of the fundamental limitations of observational studies, particularly with regard to selection bias. They and practicing clinicians were also aware of the potential negative health effects of HRT, which had to be constantly weighed against the potential cardioprotective benefits in deciding a patient's course of treatment. As a large section of the population could experience the health effects of HRT, researchers began planning RCTs to verify the promising observational study results.[71] It was highly anticipated that those RCTs would corroborate the belief that estrogen replacement can reduce CVD risk.

Randomized Controlled Trial: The Women's Health Initiative

The Women's Health Initiative (WHI) was a major study established by the National Institutes of Health in 1992 to assess a broad range of health effects in postmenopausal women. The trial was intended to follow these women for 8 years, at a cost of millions of dollars in federal funding. Among its many facets, it included an RCT to confirm the results from the observational studies discussed above. To fully investigate earlier findings, the WHI had two subgroups. One subgroup consisted of women with prior hysterectomies; they received estrogen monotherapy. The other consisted of women who had not undergone hysterectomy; they received estrogen in combination with progestin. The WHI enrolled 27,347 women in their HRT investigation: 10,739 in the estrogen-alone arm and 16,608 in the estrogen plus progestin arm. Within each arm, women were randomly assigned to receive either HRT or placebo. All women in the trial were postmenopausal and aged 50-79 years; the mean age was 63.6 years (a fact that would be important in later analysis). Some participants had experienced previous CV events. The primary outcome of both subgroups was coronary heart disease (CHD), as described by nonfatal MI or death due to CHD.

The estrogen-progestin arm of the WHI was halted after a mean follow-up of 5.2 years, 3 years earlier than expected, as the HRT users in this arm were found to be at increased risk for CHD compared to those who received placebo. The study also noted elevated rates of breast cancer and stroke, among other poor outcomes.[72,73] The estrogen-alone arm

continued for an average follow-up of 6.8 years before being similarly discontinued ahead of schedule. Although this part of the study did not find an increased risk of CHD, it also did not find any cardioprotective effect.[74] Beyond failing to locate any clear CV benefits, the WHI also found real evidence of harm, including increased risk of blood clots, breast cancer and stroke. Initial WHI publications therefore recommended against HRT being prescribed for the secondary prevention of CVD.

Where to go from here?

Researchers, and the clinicians who relied on their data for guidance in treating patients, were faced with conflicting data: epidemiological and observational studies suggested that HRT was cardioprotective while the higher-quality evidence from RCTs strongly suggested the opposite. Clinicians primarily followed the WHI results, so prescriptions for HRT in postmenopausal women quickly declined.[75] Meanwhile, researchers began to analyze the studies for potential discrepancies, and found that the women being followed in the NHS and the WHI differed in several important characteristics.

First, the WHI population was older than the NHS cohort, and many had entered menopause at least 10 years before they enrolled in the RCT. Therefore, the WHI enrollees experienced a long duration from the onset of menopause to the commencement of HRT. Meanwhile, many in the NHS population were closer to the onset of menopause and were still displaying hormonal symptoms when they began HRT.[76]

Second, although the NHS researchers adjusted the data for various confounding effects, their results could still have been subject to bias. In general, the NHS cohort was more highly educated and of a higher socioeconomic status than the WHI participants, and therefore more likely to see a physician regularly. The NHS women were also leaner and generally healthier than their RCT counterparts, and had been selected for their evident lack of pre-existing CV conditions. This selection bias in the NHS enrollment may have led to a "healthy woman" effect that in turn led to an overestimation of the benefits of therapy in the observational study.[77]

Third, researchers noted that dosing differences between the two study types may have contributed to the divergent results. The NHS reported beneficial results following low-dose estrogen therapy. The WHI, meanwhile, used a higher estrogen dose, exposing women to a larger dosage of hormones and increasing their risk for adverse events. The increased risk profile of the WHI women (e.g., older, more comorbidities, higher estrogen dose) could have contributed to the evidence of harm seen in the WHI results.[78]

Emerging Data

In addition to identifying the inherent differences between the two study populations, researchers began a secondary analysis of the NHS and WHI trials. NHS researchers

reported that women who began HRT close to the onset of menopause had a significantly reduced risk of CHD. In the subgroups of women that were older and had a similar duration after menopause compared with the WHI women, they found no significant relationship between HRT and CHD.[79] Meanwhile, WHI researchers further stratified their results by age, and found that women who began HRT close to their onset of menopause experienced some cardioprotection, while women who were further from the onset of menopause had a slightly elevated risk for CHD.[80]

Secondary analysis of both studies was therefore necessary to show that age and a short duration from the onset of menopause are crucial to HRT success as a cardioprotective agent. Neither study type provided "truth" or rather, both studies provided "truth" if viewed carefully (e.g., both produced valid and important results). The differences seen in the studies were rooted in the timing of HRT and the populations being studied.[81]

Lessons Learned From This Case Study

Although RCTs are given a higher evidence grade, observational studies provide important clinical insights. In this example, the study populations differed. For policymakers and clinicians, it is crucial to examine whether the CER was based upon patients similar to those being considered. Any study with a dissimilar population may provide non-relevant results. Thus, readers of CER need to carefully examine the generalizability of the findings being reported.

Observational Studies: Tips for the CER Consumer

- Different study types can offer different understandings; neither should be discounted without closer examination.
- RCTs provide an accurate understanding of the effect of a particular intervention in a well-defined patient group under "controlled" circumstances.
- Observational studies provide an understanding of real-world care and its impact, but can be biased due to uncontrolled factors.
- Observational studies differ in the types of databases used. These databases may lack clinical detail and contain incomplete or inaccurate data.
- Before accepting the findings from an observational study, consider whether confounding factors may have influenced the results.
- In this scenario, subgroup analysis was vital in clarifying both study designs; what is true for the many (e.g., overall, estrogen appeared to be detrimental) may not be true for the few (e.g., that for the younger post-menopausal woman, the benefits were greater and the harms less frequent).
- Carefully examine the generalizability of the study. Do the study's patients and intervention match those under consideration?
- Observational studies can identify associations but cannot prove cause-and-effect relationships.
- The Grace Principles (www.graceprinciples.org) are an excellent source of further information about observational studies and their application in comparative effectiveness research.

CHAPTER 5: DISCUSSION

The increased focus on CER should lead to a desirable expansion of evidence on what works and what does not work in health care. Much of this evidence will directly apply to patients in routine care. However, one should read CER reports carefully to ensure that what is reported is applicable, complete, and correct. This monograph presents a framework for reviewing this new CER evidence.

CER is typically performed using one of three study designs: a new clinical trial with relevant comparators; a meta-analysis, which synthesizes results across multiple trials; or an analysis of observational databases where patients received different treatments or types of care. Each of these research designs can answer certain questions, although each has limitations. Notwithstanding the differences among those designs, three general questions can be asked to determine the applicability of the research to your needs. This section will review those questions and discuss the lessons learned from the three case studies.

STEP 1: Consider for whom the findings are applicable.

Typically, reviewers consider this step *after* they determine whether the study has internal validity or had been performed with adequate rigor. However, it may be more efficient to first determine whether the results are relevant to your own needs or study question. If the research used a population, intervention, or outcomes that differ from your clinical or policy question, there is little need to analyze the study further.

The HRT case study provides a relevant example. The clinical question was: should menopausal women consider using HRT? Unfortunately, two pieces of evidence appeared to conflict: the observational NHS supported its benefits, whereas the initial WHI results did not. At the time the WHI was designed, there was an important unanswered question about HRT's potential to reduce heart disease in *older* women. The WHI showed that HRT was not effective for this use. Further, the findings that HRT increased CVD risk led to a dramatic reduction in the therapy's use. However, the WHI was not designed to answer whether HRT was safe for *younger* women of perimenopausal age to treat their menopausal symptoms—yet many applied those results too broadly. While the WHI examined a population of predominantly older women (some of whom had shown previous CV events and were therefore at increased risk), the findings were quite different when the younger women in that study were analyzed separately. For the younger, perimenopausal women, those initial conclusions were misleading as the risks were much lower in the younger subgroup. It is important to determine whether the population was relevant before adopting the results.

The meta-analysis case study presents another example of when evidence cannot be generalized. In that example, the majority of the studies in the meta-analysis had durations of one year or less. The statistical results were therefore mainly drawn from short-term events. In considering whether rosiglitazone is appropriate, the clinician is left with little concrete data about the long-term effects of the drug. The reader needs to generalize cautiously from this study if the policy question relates to long-term impact.

The cetuximab case study demonstrates the potential importance of later subgroup analysis. The original analysis showed modest benefits for cetuximab when the entire study population was analyzed. The ability of cetuximab to improve mean survival time and progression-free disease was far more apparent when the therapy was confined to those most likely to benefit because of their genetics.

All three case studies showed that the reader should remain cautious. To assume that the CER conclusions apply to your situation may be inaccurate. Determine first whether your population, intervention, environment, and outcomes are similar enough to the comparison study.

STEP 2: Consider whether any aspects of the study design might affect the results.

Once the reader judges a study to have relevance, its internal validity should be assessed. Generally, the publishing of an article indicates that experts in the field have reviewed it and have approved its quality. Nonetheless, the reader should consider possible issues with the underlying methodology, as they may affect the results. Chapter 1 reviews many such issues for each of the three CER study types. The case studies provide selected examples.

The rosiglitazone meta-analysis demonstrates that the authors' decisions in designing and conducting research may impact their findings. The reader should closely examine the study methods section of a meta-analysis report, which will reveal the choices the researchers made in carrying out their study. As seen in the rosiglitazone case, the exclusion of zero-event trials increased the weight of the CVD event rates in the other trials. Second, the studies analyzed in a meta-analysis have to be similar. It may not be suitable to combine data (e.g., outcomes) across studies if they differ in important ways, as happened with the rosiglitazone meta-analysis. Finally, the statistics used to analyze a merged dataset must be appropriate to the data being analyzed, and in that meta-analysis critics differed with the authors' choices. These "decisions" resulted in criticism of the methodology, negative impact on patient care, and reduction in rosiglitazone sales.

Observational studies are prone to bias and confounding as discussed in the HRT example. The NHS was criticized for exhibiting "selection bias" in that it followed a relatively educated cohort of women who perhaps exhibited healthier-than-average behaviors. This bias led to HRT appearing more beneficial than a controlled environment would likely show. A good CER study report should include a discussion of potential biases and how the authors accounted for them. Further, the reader should think critically about other confounders that may be present in the study that could have influenced results.

Through randomization, RCTs reduce bias and confounding. However, other problems include lack of adherence to treatment, loss to follow-up, or crossing over from one therapy group to another. Investigators should carefully assess these issues, and the analysis may adjust for them. Readers should review the authors' attempts to account for such limitations. Peer-reviewed publication generally reassures the reader that the methodologies have been examined. As the case studies showed, further critique is still warranted.

STEP 3: Consider whether the findings may change with new research.

CER may fully address a clinical question with methodologic rigor in a relevant population. However, research findings are a moving target and data may continue to emerge that clarify, modify, or reject previous conclusions. Clinical trials generally do not provide evidence about the long-term impact of an intervention. In addition, the relatively modest size of those trials means that rare events may become apparent only after broader dissemination of the treatment.

In the HRT case study, the impact of HRT initially seemed beneficial from the NHS results, then harmful from the WHI core results; a more balanced understanding was finally obtained from further analysis of the WHI. Similarly, in the case of cetuximab, advances in genetic research allowed researchers to re-examine stored biospecimens to clarify results. A genetically defined subpopulation demonstrated much greater benefit than was suggested by the original research. In the case of the rosiglitazone meta-analysis, subsequent clinical trials eroded the authors' original findings. In all of these cases, the evidentiary story evolved. As the saying goes, "We sure thought we were right then, didn't we?" This perspective should be kept in mind as standards of care are defined and access to interventions is determined based upon CER.

Conclusion

The increase in comparative effectiveness research will present the reader with an exciting array of new data to help inform health care decisions. While CER provides information, the informed reader should examine it carefully and apply the findings with caution. As solid as those findings may appear, all studies have limitations. Important criticisms may arise, further analyses may refine the results, and the study may not be adequately generalizable to your environment. The goal of new public investment in CER is to improve health outcomes and to provide additional information to the medical community so that it will better understand what works and what does not work in clinical care. The informed consumer will have multiple sources of information from which to choose for evaluating myriad health options. Education as to how to evaluate this information will be critical if it is to be used effectively to achieve the goals described.

REFERENCES:

1. IOM (Institute of Medicine). 2009. *Initial National Priorities for Comparative Effectiveness Research.* Washington, DC: The National Academies Press.

2. Pollack A. The Minimal Impact of a Big Hypertension Study. *New York Times*. 2008 Nov 27.

3. Major outcomes in high-risk hypertensive patients randomized to angiotensin-converting enzyme inhibitor or calcium channel blocker vs. diuretic: The Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT). *JAMA* 2002; 288(23):2981-2997.

4. Lieberman JA, Stroup TS, McEvoy JP, et al. Effectiveness of antipsychotic drugs in patients with chronic schizophrenia. *N Engl J Med* 2005; 353(12):1209-1223.

5. Swartz MS, Stroup TS, McEvoy JP, et al. What CATIE found: results from the schizophrenia trial. *Psychiatr Serv* 2008; 59(5):500-506.

6. Serruys PW, Morice MC, Kappetein AP, et al. Percutaneous coronary intervention versus coronary-artery bypass grafting for severe coronary artery disease. *N Engl J Med* 2009; 360(10):961-972.

7. Moser M. Results of ALLHAT: is this the final answer regarding initial antihypertensive drug therapy? *Arch Intern Med* 2003; 163(11):1269-1273.

8. Frank RG. Policy toward second-generation antipsychotic drugs: a cautionary note. *Psychiatr Serv* 2008; 59(5):521-522.

9. Tricoci P, Allen JM, Kramer JM, Califf RM, Smith SC, Jr. Scientific evidence underlying the ACC/AHA clinical practice guidelines. *JAMA* 2009; 301(8):831-841.

10. American Cancer Society. Colorectal Cancer Facts & Figures 2008-2010. 2008. Atlanta.

11. SEER Stat Fact Sheets. Surveillance Epidemiology and End Results. Available at: http://seer.cancer.gov/statfacts/html/colorect.html. Accessed September 15, 2009.

12. Rodriguez J, Viudez A, Ponz-Sarvise M, et al. Improving disease control in advanced colorectal cancer: Panitumumab and cetuximab. *Crit Rev Oncol Hematol* 2009.

13. Cunningham D, Humblet Y, Siena S, et al. Cetuximab monotherapy and cetuximab plus irinotecan in irinotecan-refractory metastatic colorectal cancer. *N Engl J Med* 2004; 351(4):337-345.

14. FDA Approval for Cetuximab. National Cancer Institute. Available at: http://www.cancer.gov/cancertopics/druginfo/fda-cetuximab. Accessed September 15, 2009.

15. Jonker DJ, O'Callaghan CJ, Karapetis CS, et al. Cetuximab for the treatment of colorectal cancer. *N Engl J Med* 2007; 357(20):2040-2048.

16. Karapetis CS, Khambata-Ford S, Jonker DJ, et al. K-ras mutations and benefit from cetuximab in advanced colorectal cancer. *N Engl J Med* 2008; 359(17):1757-1765.

17. FDA. About FDA: cetuximab (Erbitux) and panitumumab (Vectibix). Fda.gov. Available at: http://www.fda.gov/AboutFDA/CentersOffices/CDER/ucm172905.htm. Accessed September 15, 2009.

18. Normanno N, Tejpar S, Morgillo F, et al. Implications for KRAS status and EGFR-targeted therapies in metastatic CRC. *Nat Rev Clin Oncol* 2009; 6(9):519-527.

19. Allegra CJ, Jessup JM, Somerfield MR, et al. American Society of Clinical Oncology provisional clinical opinion: testing for KRAS gene mutations in patients with metastatic colorectal carcinoma to predict response to anti-epidermal growth factor receptor monoclonal antibody therapy. *J Clin Oncol* 2009; 27(12):2091-2096.

20. Frattini M, Saletti P, Romagnani E, et al. PTEN loss of expression predicts cetuximab efficacy in metastatic colorectal cancer patients. *Br J Cancer* 2007; 97(8):1139-1145.

21. Kohne CH, Lenz HJ. Chemotherapy with targeted agents for the treatment of metastatic colorectal cancer. *Oncologist* 2009; 14(5):478-488.

22. Li C, Iida M, Dunn EF, Ghia AJ, Wheeler DL. Nuclear EGFR contributes to acquired resistance to cetuximab. *Oncogene* 2009.

23. Rosell R, Moran T, Queralt C, et al. Screening for epidermal growth factor receptor mutations in lung cancer. *N Engl J Med* 2009; 361(10):958-967.

24. American Heart Association. Heart Attack and Angina Statistics. American Heart Association. Available at: http://www.americanheart.org/presenter.jhtml?identifier=4591. Accessed October 19, 2009.

25. Centers for Disease Control and Prevention. National diabetes fact sheet: general information and national estimates on diabetes in the United States, 2007. 2008. Atlanta, GA, U.S. Department of Health and Human Services, Centers for Disease Control and Prevention.

26. National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK). National Diabetes Information Clearinghouse. NIH Publication No.09-3873. Available at: http://diabetes.niddk.nih.gov/dm/pubs/overview/index.htm. Accessed August 5, 2009.

27. Economic costs of diabetes in the U.S. In 2007. *Diabetes Care* 2008; 31(3):596-615.

28. Wieczorek A, Rys P, Skrzekowska-Baran I, Malecki M. The role of surrogate endpoints in the evaluation of efficacy and safety of therapeutic interventions in diabetes mellitus. *Rev Diabet Stud* 2008; 5(3):128-135.

29. Meinert CL, Knatterud GL, Prout TE, Klimt CR. A study of the effects of hypoglycemic agents on vascular complications in patients with adult-onset diabetes. II. Mortality results. *Diabetes* 1970; 19:Suppl-830.

30. Leibel B. An analysis of the University Group Diabetes Study Program: data results and conclusions. *Can Med Assoc J* 1971; 105(3):292-294.

31. Matthews DR, Tsapas A. Four decades of uncertainty: landmark trials in glycaemic control and cardiovascular outcome in type 2 diabetes. *Diab Vasc Dis Res* 2008; 5(3):216-218.

32. Phillips LS, Grunberger G, Miller E, et al. Once- and twice-daily dosing with rosiglitazone improves glycemic control in patients with type 2 diabetes. *Diabetes Care* 2001; 24(2):308-315.

33. Lebovitz HE, Dole JF, Patwardhan R, Rappaport EB, Freed MI. Rosiglitazone monotherapy is effective in patients with type 2 diabetes. *J Clin Endocrinol Metab* 2001; 86(1):280-288.

34. Fonseca V, Rosenstock J, Patwardhan R, Salzman A. Effect of metformin and rosiglitazone combination therapy in patients with type 2 diabetes mellitus: a randomized controlled trial. *JAMA* 2000; 283(13):1695-1702.

35. Sutton MStJ, Rendell M, Dandona P, et al. A comparison of the effects of rosiglitazone and glyburide on cardiovascular function and glycemic control in patients with type 2 diabetes. *Diabetes Care* 2002; 25(11):2058-2064.

36. Raskin P, Rendell M, Riddle MC, et al. A randomized trial of rosiglitazone therapy in patients with inadequately controlled insulin-treated type 2 diabetes. *Diabetes Care* 2001; 24(7):1226-1232.

37. Gerstein HC, Yusuf S, Bosch J, et al. Effect of rosiglitazone on the frequency of diabetes in patients with impaired glucose tolerance or impaired fasting glucose: a randomised controlled trial. *Lancet* 2006; 368(9541):1096-1105.

38. Mulrow CD, Cornell J, Localio AR. Rosiglitazone: a thunderstorm from scarce and fragile data. *Ann Intern Med* 2007; 147(8):585-587.

39. Kahn SE, Haffner SM, Heise MA, et al. Glycemic durability of rosiglitazone, metformin, or glyburide monotherapy. *N Engl J Med* 2006; 355(23):2427-2443.

40. Nissen SE, Wolski K. Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes. *N Engl J Med* 2007; 356(24):2457-2471.

41. The Cochrane Handbook for Systematic Reviews of Interventions 4.2.6. The Cochrane Collaboration. Available at: http://www.cochrane.org/resources/handbook. Accessed August 5, 2009.

42. FDA News Release: FDA Issues Safety Alert on Avandia. U.S.Food and Drug Administration. Available at: http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/2007/ucm108917.htm. Accessed July 16, 2009.

43. Starner CI, Schafer JA, Heaton AH, Gleason PP. Rosiglitazone and pioglitazone utilization from January 2007 through May 2008 associated with five risk-warning events. *J Manag Care Pharm* 2008; 14(6):523-531.

44. Surowiecki J. A Drug on the Market. The New Yorker. Available at: http://www.newyorker.com/talk/financial/2007/06/25/070625ta_talk_surowiecki. Accessed July 27, 2009.

45. Bloomgarden ZT. The Avandia debate. *Diabetes Care* 2007; 30(9):2401-2408.

46. Diamond GA, Bax L, Kaul S. Uncertain effects of rosiglitazone on the risk for myocardial infarction and cardiovascular death. *Ann Intern Med* 2007; 147(8):578-581.

47. Goldfine AB. The rough road for rosiglitazone. *Curr Opin Endocrinol Diabetes Obes* 2008; 15(2):113-117.

48. Misbin RI. Lessons from the Avandia controversy: a new paradigm for the development of drugs to treat type 2 diabetes. *Diabetes Care* 2007; 30(12):3141-3144.

49. Psaty BM, Furberg CD. Rosiglitazone and cardiovascular risk. *N Engl J Med* 2007; 356(24):2522-2524.

50. Rosen CJ. The rosiglitazone story--lessons from an FDA Advisory Committee meeting. *N Engl J Med* 2007; 357(9):844-846.

51. Ajjan RA, Grant PJ. The cardiovascular safety of rosiglitazone. *Expert Opin Drug Saf* 2008; 7(4):367-376.

52. Mannucci E, Monami M, Marchionni N. Rosiglitazone and cardiovascular risk. *N Engl J Med* 2007; 357(9):938-940.

53. Friedrich JO, Adhikari NK, Beyene J. Inclusion of zero total event trials in meta-analyses maintains analytic consistency and incorporates all available data. *BMC Med Res Methodo*l 2007; 7:5.

54. Hulley SB, Cummings SR, Browner WS, et al. *Designing Clinical Research.* 2nd ed. Philadelphia, PA: Lippincott Williams & Wilkins; 2001.

55. Dahabreh IJ, Economopoulos K. Meta-analysis of rare events: an update and sensitivity analysis of cardiovascular events in randomized trials of rosiglitazone. *Clin Trials* 2008; 5(2):116-120.

56. Bracken MB. Rosiglitazone and cardiovascular risk. *N Engl J Med* 2007; 357(9):937-938.

57. Diamond GA, Kaul S. Rosiglitazone and cardiovascular risk. *N Engl J Med* 2007; 357(9):938-939.

58. Home PD, Pocock SJ, Beck-Nielsen H, et al. Rosiglitazone evaluated for cardiovascular outcomes--an interim analysis. *N Engl J Med* 2007; 357(1):28-38.

59. Home PD, Pocock SJ, Beck-Nielsen H, et al. Rosiglitazone evaluated for cardiovascular outcomes in oral agent combination therapy for type 2 diabetes (RECORD): a multicentre, randomised, open-label trial. *Lancet* 2009; 373(9681):2125-2135.

60. Singh S, Loke YK, Furberg CD. Long-term risk of cardiovascular events with rosiglitazone: a meta-analysis. *JAMA* 2007; 298(10):1189-1195.

61. Framingham Heart Study. National Heart, Lung and Blood Institute and Boston University. Available at: http://www.framinghamheartstudy.org/. Accessed September 18, 2009.

62. Lloyd-Jones D, Adams R, Carnethon M, et al. Heart disease and stroke statistics--2009 update: a report from the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. *Circulation* 2009; 119(3):480-486.

63. Grodstein F, Stampfer M. The epidemiology of coronary heart disease and estrogen replacement in postmenopausal women. *Prog Cardiovasc Dis* 1995; 38(3):199-210.

64. Stevenson JC. HRT and cardiovascular disease. *Best Pract Res Clin Obstet Gynaecol* 2009; 23(1):109-120.

65. Barrett-Connor E, Grady D. Hormone replacement therapy, heart disease, and other considerations. *Annu Rev Public Health* 1998; 19:55-72.

66. Krieger N, Lowy I, Aronowitz R, et al. Hormone replacement therapy, cancer, controversies, and women's health: historical, epidemiological, biological, clinical, and advocacy perspectives. *J Epidemiol Community Health* 2005; 59(9):740-748.

67. Stampfer MJ, Colditz GA, Willett WC, et al. Postmenopausal estrogen therapy and cardiovascular disease. Ten-year follow-up from the nurses' health study. *N Engl J Med* 1991; 325(11):756-762.

68. Grodstein F, Stampfer MJ, Manson JE, et al. Postmenopausal estrogen and progestin use and the risk of cardiovascular disease. *N Engl J Med* 1996; 335(7):453-461.

69. Grodstein F, Manson JE, Colditz GA, et al. A prospective, observational study of postmenopausal hormone therapy and primary prevention of cardiovascular disease. *Ann Intern Med* 2000; 133(12):933-941.

70. Varas-Lorenzo C, Garcia-Rodriguez LA, Perez-Gutthann S, Duque-Oliart A. Hormone replacement therapy and incidence of acute myocardial infarction. A population-based nested case-control study. *Circulation* 2000; 101(22):2572-2578.

71. Hulley S, Grady D, Bush T, et al. Randomized trial of estrogen plus progestin for secondary prevention of coronary heart disease in postmenopausal women. Heart and Estrogen/progestin Replacement Study (HERS) Research Group. *JAMA* 1998; 280(7):605-613.

72. Rossouw JE, Anderson GL, Prentice RL, et al. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results From the Women's Health Initiative randomized controlled trial. *JAMA* 2002; 288(3):321-333.

73. Manson JE, Hsia J, Johnson KC, et al. Estrogen plus progestin and the risk of coronary heart disease. *N Engl J Med* 2003; 349(6):523-534.

74. Anderson GL, Limacher M, Assaf AR, et al. Effects of conjugated equine estrogen in postmenopausal women with hysterectomy: the Women's Health Initiative randomized controlled trial. *JAMA* 2004; 291(14):1701-1712.

75. Hersh AL, Stefanick ML, Stafford RS. National use of postmenopausal hormone therapy: annual trends and response to recent evidence. *JAMA* 2004; 291(1):47-53.

76. Rosano GM, Vitale C, Lello S. Postmenopausal hormone therapy: lessons from observational and randomized studies. *Endocrine* 2004; 24(3):251-254.

77. Posthuma WF, Westendorp RG, Vandenbroucke JP. Cardioprotective effect of hormone replacement therapy in postmenopausal women: is the evidence biased? *BMJ* 1994; 308(6939):1268-1269.

78. Radford N, Church T. Risks of postmenopausal hormone replacement. *JAMA* 2002; 288(22):2819-2824.

79. Grodstein F, Manson JE, Stampfer MJ. Hormone therapy and coronary heart disease: the role of time since menopause and age at hormone initiation. *J Womens Health (Larchmt)* 2006; 15(1):35-44.

80. Rossouw JE, Prentice RL, Manson JE, et al. Postmenopausal hormone therapy and risk of cardiovascular disease by age and years since menopause. *JAMA* 2007; 297(13):1465-1477.

81. Vandenbroucke JP. The HRT controversy: observational studies and RCTs fall in line. *Lancet* 2009; 373(9671):1233-1235.

1894 Preston White Drive
Reston, VA 20191-5433

1501 M Street, NW
Suite 650
Washington, DC 20005

Phone: 703.620.6390
Fax: 703.476.0904

www.npcnow.org