**Homework Problem 3**: Form a group of 2 or 3 students (there should be two groups of 2 and one group of 3 students). Each group should pick one of the three case-studies below. Please avoid overlaps – first come-first serve. Coordinate groups and group case-study choices. The group should review the case-study, go deeper in to the provenance of the study, review study goals, design, data, approaches and findings and give an in-class presentation (30-min + Q&A). Try to frame the case-study in methodological terms synergizing with some of the techniques we discuss in class. Although there are significant details about each case-study, you are encouraged to research the study further. Each group will submit their slides and report (all members independently submit the same files on Canvas) and present the report in class (around the due date). Examples of themes to cover in your report include:

* Why was this study undertaken? Summarize the goals.
* What were the study design, data and approach?
* What are the study outcomes, clinically-relevant contributions, and healthcare impact?
* In retrospect, are there aspects of the study that could have been done differently?
* How can the study be extended? What would you do next?

1. **Case-Study 1: The Cetuximab Study [[1]](#footnote-1)**

**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 CRTs examined its impact in patients with mCRC (metastatic-stage Colorectal cancer). In the first one, 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. The study reported 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). As 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.

A second CRT (2007) examined 572 patients and suggested efficacy of cetuximab in the treatment of mCRC. This 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. It reported 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?**

These RCTs 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 epidermal growth factor receptor (EGFR) treatment mechanism (Cetuximab,) 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. 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 4 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.

**What next?**

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. This distinction reduces the relevant population by approximately 40%. 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. 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; and similar genetic targeting is being investigated, and advocated, in other cancer types.

**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 suggests that cetuximab initially appeared to have rather modest clinical benefits. Albeit, new information that 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. When improperly employed, post hoc analyses may lead to incorrect patient care conclusions.

**RCTs Tips for the CER Practitioners**

* + 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.

1. **Case-Study 2: The Rosiglitazone Study [[2]](#footnote-2)**

**Meta-analysis**

Often the 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.

**Clinical Application**

Heart disease is the leading cause of mortality in the United States, resulting in approximately 20% of all deaths. Diabetics are particularly susceptible to heart disease, with more than 65% of deaths attributable to it. The nonfatal complications of diabetes are wide-ranging and include kidney failure, nerve damage, amputation, stroke and blindness, among other outcomes. In 2007, the total estimated cost of diabetes in the United States was $174B; $116B was derived from direct medical expenditures and the rest from the indirect cost of lost productivity due to the disease. 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.

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. 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). Since these early studies, CV concerns continue to be an issue with selected oral hypoglycemic agents that have subsequently entered the marketplace.

A class of drugs, thiazolidinedione (TZD), was approved in the late 1990s, as 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. However, safety concerns developed as the FDA received reports of adverse cardiac events potentially associated with rosiglitazone. 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.

**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 study began with a broad search to locate potential studies for review. The authors 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. 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. The 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. Analyzing their data, the authors chose one particular statistical method (the Peto odds ratio method, a fixed-effect statistical approach), which calculates the odds of events occurring where the outcomes of interest are rare and small in number. 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. These findings appeared online on the same day that the FDA issued a safety alert regarding rosiglitazone. Discussion of the meta-analysis was immediately featured prominently in the news media. By December 2007, prescription claims for the drug at retail pharmacies had fallen by more than 50%.

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. The 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, and for combining both placebo and drug therapy data into one comparator arm. 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. 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. 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. The researchers accepted studies and populations that were clinically heterogeneous, yet pooled them as if they were not. The study reported 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 critiques focused on the authors’ use of the Peto method as being an incorrect choice because data were pooled from both small and very large studies, resulting in a potential overestimation of treatment effect. Others reviewers pointed 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. 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. 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. 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. 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. 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. 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.

**Now what?**

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. 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. These 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 CER Practitioners**

* 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.

1. **Case-Study 3: The Nurses’ Health Study [[3]](#footnote-3)**

**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. Such 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. 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.

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. Critical assessment of whether the described results are helpful or biased (based upon how the study was performed) are necessary. This case will illustrate several characteristics of the types of studies that will assist in evaluating newly published work.

**Clinical Applications**

Cardiovascular diseases (CVD) are the leading cause of death in women older than the age of 50. 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. 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.

**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. Later, estrogen-progestin combination therapy was shown to be even more cardioprotective than estrogen monotherapy, and lower doses of estrogen replacement therapy were found to deliver equal cardioprotection and lower the risk for adverse events. 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.

**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 secondary retrospective data analysis of the UK primary care electronic medical record database, for example, also showed the protective effect associated with HRT use. 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. 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 second group 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. 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. 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.

**What Next?**

Scientists 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. 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. Thus, the WHI enrollees experienced a long duration from the onset of menopause to the commencement of HRT. At the same time, many in the NHS population were closer to the onset of menopause and were still displaying hormonal symptoms when they began HRT. 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. 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 WHL, 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.

**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. Also, the WHI study further stratified these 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.

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.

**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 Practitioners**

• 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.

1. <http://www.cancer.gov/cancertopics/druginfo/fda-cetuximab> [↑](#footnote-ref-1)
2. <http://www.nejm.org/doi/full/10.1056/NEJMoa072761> [↑](#footnote-ref-2)
3. <http://jech.bmj.com/content/59/9/740.short> [↑](#footnote-ref-3)