HORMONE REPLACEMENT THERAPY AND WOMEN'S HEALTH: DO THE LATEST STUDIES JUSTIFY MAJOR CONCERNS?

Phil Clapham

This is a lay review of the following paper:

Women's Health Initiative. 2002. Risks and benefits of estrogen plus progestin in healthy menopausal women. Principal results from the Women's Health Initiative Randomized Controlled Trial. *Journal of the American Medical Association* 288: 321-333.

My comments below are not intended to provide medical advice, but rather to give women with an interest in this controversial issue some information about the work on which recent claims of concern have been based. This is a personal scientific perspective which tries to clarify the issues involved.

Bottom line

This is a study about the effects of hormone replacement therapy. The study is flawed, and its results have been misrepresented in the press or cited as alarming without appropriate caveats. My conclusion after reviewing this work is as follows:

If a woman is taking hormones (estrogen plus progestin) in the hope of reducing risk of secondary medical problems such as heart disease, she should probably stop since - as the study says - there's no clear benefit (and never has been, despite an earlier - also flawed - study suggesting an effect). But if a woman is using hormone replacement to alleviate unpleasant symptoms associated with menopause, there's no good evidence in this study to discontinue hormone use since the increased risk, if any, is small in absolute terms.

Introduction and overview

This paper summarizes the results of a large study (a component of the Women's Health Initiative) which sought to assess the major health benefits and risks of the most commonly used form of hormone replacement therapy (HRT) in the United States. This form is the combination of estrogen plus progestin, marketed by Wyeth Corporation. In May of this year the study examined the results of several years of follow-up (mean 5.2 years) and recommended discontinuing the trial. This decision was based upon the conclusion that the risk of invasive breast cancer (and to a lesser extent other diseases) exceeded any benefits derived from HRT.

The paper has been widely cited in the press and has resulted in much discussion and anguish among women using HRT. Because the work is clearly influencing national health policy, it is

worthwhile to take a detailed look at how the research was conducted and whether its conclusions are justified. This is the aim of this informal review.

Objective. The major objective of the study was to see whether the most commonly used hormone treatment provided any secondary health benefit or conferred greater risks to users. Prior to this analysis, an initial study of HRT had suggested that hormones could help to reduce to risk of coronary heart disease, as well as reducing loss of bone density and other problems. The study does not address the fact that many younger women use HRT not for secondary health benefits, but to relieve symptoms associated with menopause such as night sweats, hot flashes and depression.

Sample size. The study is based upon a large sample size: 16,608 post-menopausal women. These were divided into two groups: one given hormones (8,506 women) and the other given a placebo (8,102). The groups were randomized with respect to demographic, life history and medical variables, and the study was conducted blind. As I note below, there are some serious problems with the way in which the analyses of these two groups were conducted.
What was looked at. Essentially the study examined the rates at which the women in the two groups (HRT and placebo) were diagnosed with various medical problems over the study period, and attempted to examine whether these differences were statistically significant. The medical problems included various types of cardiovascular disease, cancer and fractures.

Statistical analysis. This entire study is based upon proportional hazards analysis. What this amounts to at base is taking the number of instances of each reported medical problem in each group and calculating the ratio for each problem between the two groups, adjusted for sample size. To give a simplistic example for a hypothetical study involving equal sample sizes, 6 occurrences of a disease in one group and 4 in the other would yield a Hazard Ratio of 1.50. While this oversimplifies what was done in the HRT study, that's the basis of the analysis.

Major results. The study gives Hazard Ratios (HR) with 95% confidence intervals which - the authors claim - suggest that hormone therapy put the women in the HRT group at higher risk of incurring various medical problems than those in the placebo group. These problems include heart disease, breast cancer and stroke. The study also claims that HRT reduces the risk of colorectal cancer, endometrial cancer, and hip fractures. The major conclusion was that HRT should not be used as an intervention to help prevent heart disease (contradicting the results of the earlier study).

Critique

This study has a potentially huge flaw which makes most of my other comments moot. But before I discuss the flaw, here are some other notes and problems.

First of all, it is probably worth noting that the concept of hypothesis testing appears to be foreign to this study (and to much of the medical profession generally). It isn't at all clear why the

questions being addressed were asked in the first place. This may seem churlish, but conducting a study to examine the effect of HRT (or anything else) on a variety of unrelated health problems, without any underlying biological framework with which to explain possible cause-effect, seems like one large and vague fishing expedition. The paper does not include any mention of whether there was ever any biological basis for believing that hormone treatments would affect risk of, say, heart disease, either negatively or positively.

Perhaps this is given elsewhere; but it would be nice to see a medical researcher summarize the biological mechanisms underlying putative impact, and formulate some hypotheses about such impact accordingly. In other words, saying something like "Hormones such as these are thought to impact heart disease through such and such mechanism, therefore we hypothesize that women taking these hormones are at greater risk, and we tried to test this".

For the many women undergoing HRT, it is important to clearly articulate the questions that this study seeks to answer. If the question is whether women should be using HRT to help prevent certain diseases (such as heart disease), then the answer is pretty clearly No; there is no clear benefit to HRT in this regard. If, however, the question is whether use of HRT to reduce problematic menopausal symptoms puts *an individual woman* at substantially increased risk for certain diseases, the answer is probably not.

Looking at this in the individual context is critical, since the absolute risk increases involved - i.e. how much an individual woman increases her risk of disease through HRT - are very small. Much press attention has focused on the Hazard Ratios, which suggest that HRT users may be (e.g.) "39% more likely" to incur problem X if they continue hormone use. But relative rates are misleadingly alarmist. True, 6 cases of something is 50% more than 4 cases of that thing; but 6 versus 4 instances of *anything* in a sample of several thousand women is a meaningless difference, and translates to a very low absolute risk for an individual.

What women need to look at here is *absolute* (not relative) risk. Even if one accepts the analyses of this paper (which there are good reasons not to), the absolute risk increase from use of HRT translates as follows, expressed as excess risk per 10,000 person-years: 7 more cases of coronary heart disease, 8 more strokes, 8 more pulmonary embolisms, and 8 more invasive breast cancers. This gives a very different picture from one which posits a "39%" increase in risk.

Problems with the study's analyses

The validity of the analyses in this study is pretty much entirely contingent upon the validity of the hazard rate "model" used. Not being a statistician, I don't want to venture into this territory. But the underlying principal is straightforward: comparing relative risk rates in the way described above.

Again, these hazard ratios are based on the number of reported incidences of different health problems in the two groups (HRT versus placebo), and these numbers are small relative to the overall sample size of more than 16,000. Whether the claimed statistically significant differences in such ratios are valid is at the heart of this problem; I have a very hard time believing that any of the ratios could be significantly different. The only exception may be venous thromboembolic disease (which includes deep-vein thrombosis and pulmonary embolism); and it's noteworthy that this is the *only* health problem involved here which the authors of the study suggest was predictable based upon known biological effectors.

But here's the big problem (I'll try to make this as clear as possible). The sample size for this study, again, was 16,608 women. Of these, 8,506 were in the HRT group, and 8,102 in the placebo group. The analyses are all conducted on this *entire* sample, and those analyses are based on the assumption that all women in those groups did what they were supposed for the duration of the trial (i.e. they either took hormones, or they didn't take any). The hazard ratios given in the paper for each of the medical problems discussed, and the differences in these ratios between the HRT and placebo group, are all calculated on this assumption. With me so far?

It is rather surprising, therefore, to find on page 326 that "a substantial number of women had stopped taking study drugs at some time" or that some women in the placebo group "initiated hormone use through their own clinician". How substantial is "substantial"? Turns out it that a whopping 42% of those in the HRT group stopped taking HRT at some point, while 10.7% of those in the placebo group actually started taking hormones at some point before the end of the study. These figures are termed "crossover rates" and the authors note that the rates in both directions "exceeded design projections".

Despite this massive methodological complication, the study blithely goes on to analyze *all 16,608 women* as if nothing had happened, and to calculate relative risk between HRT and placebo accordingly.[1] Then, in one of the more twisted examples of spurious logic I've seen in a while, they actually go on to state that:

[1] There is an added complication which I won't bring up, in that the total sample also includes several hundred women who were "lost" to follow-up or stopped providing outcome information. Despite this, they're included in the analyses; however, this problem is dwarfed by the crossover issue.

> *the lack of adherence would tend to decrease the observed treatment effects. Thus, the results presented here may underestimate the magnitude of... adverse effects on cardiovascular disease and breast cancer*

The logic is that, since a lot of the women in the HRT group actually stopped taking hormones, the observed risks would have been even greater if they hadn't.

This is the scientific equivalent of having your cake and eating it too. First, you produce results concerning risks and clinical outcomes of hormone therapy based on a sample of 8,506 women, even though 42% of that sample didn't adhere to the treatment. Then you act as if the results were strictly valid for the *entire* sample and claim that they would have been even more marked if everyone had adhered to the protocol. The problem, of course, is that almost half of them didn't, so it's inappropriate to report disease rates for these women and to ascribe a connection to HRT.

The simple solution to this problem would of course be to remove all non-adhering patients from the sample (together with the several hundred women from the footnote for which complete follow-up information was not available) and then analyze what was left. This is known as *censoring* data; typically you'd use information on the women concerned only up to the point where they "crossed over". Why this elementary step was not taken is the biggest mystery in this study. The paper does not tell us whether they attempted to do this, nor what impact the smaller sample (and lower statistical power) would have had on the results.

Conclusion

This is one more example of flawed medical research setting national health policy and scaring large numbers of people. Medical science is full of examples in which the latest study contradicts previous work, and I strongly suspect that much of this is due to the poor analysis and study design involved. Until doctors stop going on giant poorly designed fishing expeditions and instead learn to formulate biologically sound hypotheses and to analyze data with appropriate rigor, their work will continue to misinterpret small effects and chance variations, to the detriment of us all.