New evidence rekindles the hormone therapy debate

Noreen Goldman

Background

In a now well-publicised series of events during 2002 and 2003, the Women’s Health Initiative (WHI) and other randomised controlled trials (RCTs) designed to examine the links between postmenopausal hormone therapy (HT) and the prevention of chronic disease were terminated prematurely. The estrogen plus progestin arm of the WHI was terminated when a global health index indicated overall harm and a significantly higher breast cancer risk for the treatment group, and the estrogen arm was stopped because of an increased risk of stroke in the absence of evidence of cardioprotection.

Expectations that HT would be beneficial for the prevention of heart disease were not borne out by the RCT results. In contrast to most observational studies that reported lower rates of coronary heart disease (CHD) for long-term users of HT, RCTs suggested no reduction or an increased incidence. In particular, estimates of the relative risk of CHD associated with HT use from the estrogen arm of the WHI suggested no cardiovascular benefit, whereas those from the estrogen plus progestin trial indicated excess risks of a magnitude of 1.2 to 1.3. These conflicting findings generated a huge response from the public, researchers and clinicians.

Recent evidence regarding the sources of discrepancy between the observational studies and the RCTs provides us with a timely opportunity for a fresh perspective on the HT controversy. These analyses, which have received far less publicity than the initial WHI findings, oblige us to both examine the implications of the exalted status of RCTs vis-à-vis observational research and consider when and how to disseminate the revised findings to women and clinicians.

The HT controversy and its consequences

In the presence of divergent findings between many well-conducted observational studies, such as the Nurses’ Health Study (NHS), and a small number of high quality RCTs, the RCTs triumphed. Epidemiologists assessing the discrepancies between the two types of studies generally implicated biases or confounders in the observational data. Some considered the HT controversy to be a debacle for epidemiological research and others went so far as to predict the demise of observational studies. Nevertheless, some researchers levelled criticisms against the RCTs, pointing out, for example, that differential rates of unblinding between treatment and control groups, high discontinuation and crossover rates, and a relatively old age distribution of participants in the trials – discussed in further detail later – could have biased estimates from the RCTs.

The consequences of the WHI findings went far beyond academic debate. The prevalence of HT plummeted in the USA and in other countries beginning in 2002; clinical guidelines were modified to advise against the use of HT for prevention of cardiovascular disease; and postmenopausal women and their physicians were confronted with difficult decisions about hormonal use. Women with unpleasant or debilitating menopausal symptoms frequently abandoned HT and many continue their struggle to find satisfactory alternatives.

Resolving the discrepancies

Various studies suggest that the discrepancies between observational studies and RCTs may be due in large part to differences in the timing of HT initiation relative to the onset of menopause. Women taking hormones in observational studies had typically begun therapy in early menopause, whereas participants in the major RCTs had been assigned treatment more than a decade after menopause, on average. Several types of evidence supporting the importance of a woman’s age or duration since menopause at the initiation of HT have now emerged: (1) a secondary analysis based on pooled data from the two WHI trials yielded a significant trend in relative risk for CHD, with women at higher durations since menopause experiencing higher relative risks than those adopting HT closer to menopause onset; (2) re-analysis of the NHS showed a similar pattern; (3) stratification of WHI and NHS estimates of CHD risk by timing of HT initiation eliminated much of the difference between the two sets of results; and (4) a meta-analysis based on 23 RCTs indicated lower CHD risk among HT users than non-users only for studies comprising relatively young postmenopausal women or women within a decade of menopause onset. Although not all trends were statistically significant, estimates of relative risk for incidence of coronary disease and total mortality derived from these studies (Table 1) show a consistent pattern by age or time since menopause: protective effects at early ages or durations of HT use and progressively higher relative risks in later years. Additional evidence comes from the WHI Coronary Artery Calcium Study (WHI-CACS), which was initiated in 2004 among 50–59-year-old women in the estrogen-only arm of WHI: estrogen users had lower levels of calcified plaque (a marker of total atherosclerotic plaque burden) in their coronary arteries compared with the placebo group.

The evidence for age modification of HT effects is strengthened by its biological plausibility. Clinical trials in women and experiments involving non-human primates suggest that exogenous estrogen has complex, counteracting effects on coronary function that vary with the stage of atherosclerosis: at early stages, estrogen may lower coronary risk by improving lipid and endothelial function, but the prothrombotic and pro-inflammatory effects of estrogen may result in clotting or rupture of plaque at more advanced stages. Thus, a woman’s age and time since menopause at HT initiation are likely to be surrogate variables for the degree of vascular disease.

Limitations of RCTs

These new findings underscore some of the well-known limitations of RCTs. Despite serving as the gold standard of...
Goldman

Table 1 Risks for coronary heart disease and total mortality associated with hormone therapy, by age and years since menopause

<table>
<thead>
<tr>
<th>Study data</th>
<th>Age (years)</th>
<th>50–59</th>
<th>60–69</th>
<th>70–79</th>
<th>p value for trend*</th>
<th>Years since menopause (n)</th>
<th>10–19</th>
<th>20+</th>
<th>p value for trend*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Women’s Health Initiative (WHI)29</td>
<td>1.29 (0.79–2.12)</td>
<td>1.03 (0.74–1.43)</td>
<td>1.48 (1.04–2.11)</td>
<td>0.70 (0.54–1.43)</td>
<td>1.23 (0.85–1.77)</td>
<td>1.66 (1.14–2.41)</td>
<td>0.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estrogen + progestin</td>
<td>0.63 (0.36–1.09)</td>
<td>0.94 (0.71–1.24)</td>
<td>1.13 (0.82–1.54)</td>
<td>0.12 (0.20–1.17)</td>
<td>0.96 (0.64–1.44)</td>
<td>1.12 (0.86–1.46)</td>
<td>0.15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estrogen alone</td>
<td>0.93 (0.65–1.33)</td>
<td>0.98 (0.79–1.21)</td>
<td>1.26 (1.00–1.59)</td>
<td>0.16 (0.50–1.16)</td>
<td>1.10 (0.84–1.45)</td>
<td>1.28 (1.03–1.58)</td>
<td>0.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Combined HT</td>
<td>&lt;60</td>
<td>0.86 (0.65–1.14)</td>
<td>1.15 (0.85–1.57)</td>
<td>– (–)</td>
<td>0.05 (0.61–1.14)</td>
<td>1.12 (0.84–1.48)</td>
<td>– (–)</td>
<td>0.08</td>
<td></td>
</tr>
<tr>
<td>Nurses’ Health Study6 (NHS)26</td>
<td>0.78 (0.44–1.90)</td>
<td>1.36 (0.81–2.29)</td>
<td>– (–)</td>
<td>0.06 (0.19–1.51)</td>
<td>1.20 (0.78–1.84)</td>
<td>– (–)</td>
<td>0.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estrogen + progestin (adherence adjusted)</td>
<td>0.68 (0.48–0.96)</td>
<td>1.03 (0.91–1.16)</td>
<td>– (–)</td>
<td>– (–)</td>
<td>– (–)</td>
<td>– (–)</td>
<td>– (–)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Meta-analysis5 (23 RCTs)31</td>
<td>0.70 (0.51–0.96)</td>
<td>1.05 (0.67–1.26)</td>
<td>1.14 (0.94–1.37)</td>
<td>0.06 (0.53–1.09)</td>
<td>0.98 (0.78–1.24)</td>
<td>1.14 (0.96–1.36)</td>
<td>0.51</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WHI (combined HT)29</td>
<td>&lt;60</td>
<td>0.61 (0.39–0.95)</td>
<td>1.03 (0.90–1.18)</td>
<td>– (–)</td>
<td>– (–)</td>
<td>– (–)</td>
<td>– (–)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The p values for trend for the WHI are based on parameterisations of age as a linear form of categorical coded values. The p values for the NHS are tests for interactions.

The meta-analysis for CHD, RCTs are classified in the two following groups: (1) mean age of participants <60 years or mean years from menopause <10 years; (2) mean age of participants >60 years or mean years from menopause >10 years. In the meta-analysis for total mortality, the two groups are defined as (1) mean age <60 years and (2) mean age >60 years.

Estimates from the meta-analyses are odds ratios.

CHD, coronary heart disease; CI, confidence interval; HT, hormone therapy; NHS, Nurses’ Health Study; RCT, randomised controlled trial; WHI, Women’s Health Initiative.

clinical research, their lack of external validity continues to be a serious concern. The sample participating in clinical trials is rarely a random sample from a well-defined population or representative of the population of interest. Although cost and sample size considerations, along with clinical interest in prescribing HT for disease prevention in older women, led WHI investigators to implement a broad age range (50–79 years), the vast majority of women outside of clinical trials adopt HT within several years of menopause. For example, 80% of hormone users in the NHS initiated use more than a decade after menopause (51 years, on average, in the USA). In contrast, NHS initiated use within a 2–3-year period after the start of menopause (<10 years; 2) mean age of participants >60 years or mean years from menopause >10 years. In the meta-analysis for total mortality, the two groups are defined as (1) mean age <60 years and (2) mean age >60 years.

Subsequent to WHI up of the WHI samples, it is unlikely that these data can ever provide a convincing answer to the question of whether HT is protective of cardiovascular disease for women who initiate HT early.

This lesson is not a new one, although it provides an intriguing reversal of a problem associated with RCTs of commonly used treatments for chronic disease. In particular, RCTs of statins and of non-steroidal anti-inflammatory drugs have been criticised for under-representation of older participants. Subsequent to US Food and Drug Administration (FDA) approval, an unexpectedly large number of serious, and sometimes fatal, side effects among the elderly taking some of these medications resulted in withdrawals of these drugs from the average effects in heterogeneous groups, rather than on variation in those effects by clinically relevant characteristics of participants. WHI was not designed or statistically powered to examine the consequences of HT for women close to the menopausal transition or to assess variation in the health consequences of HT by age or time since menopause (or stage of vascular disease), and certainly not to statistically distinguish the effects of these highly correlated variables. Indeed, in early analyses, WHI investigators did not find statistically significant differences in HT effects by age or other covariates. Thus, despite informative post hoc analyses and ongoing follow up of the WHI samples, it is unlikely that these data can ever provide a convincing answer to the question of whether HT is protective of cardiovascular disease for women who initiate HT early.
market. In contrast, the major external validity problem for RCTs of HT is under-representation of relatively young (postmenopausal) women rather than the elderly.

Although recent analyses of discrepancies between the RCTs and observational studies have focused much of their attention on disparities in the underlying ages and durations since menopause of study participants, additional WHI limitations may have given rise to the conflicting findings. For example, the much higher proportions of women unblinded in the estrogen–progestin arm of the WHI (primarily because of vaginal bleeding) as compared with the placebo group (0.44 vs 0.07) made the trial resemble an observational study. Detection bias resulting from this differential awareness of HT status between treatment and control participants may have led to overestimates of the relative risks of several health conditions, including CHD.

Disseminating new findings and revising clinical guidelines

Another salient issue emanating from the HT controversy pertains to dissemination of the new findings. Despite numerous published studies that identify the importance of age at HT initiation, many women would be hard pressed to locate information challenging the 2002 WHI results or discussing variation in risk. Most of the frequently accessed health websites do not acknowledge the lack of evidence linking HT to increased risk of CHD for women initiating HT in the usual manner. Nor do they provide information suggesting that HT use by these women may confer protection against CHD for short to moderate durations of use. Researchers are undoubtedly concerned about widespread dissemination of their “revisionist” thinking in the absence of new RCT data and conclusive results, and face the difficulty of communicating such complex findings to a broad audience.

A similar problem relates to potential revision of clinical guidelines for HT use. Neither positive findings from the WHI-CACS study nor re-analyses of WHI and NHS data have generated recommendations for changing existing guidelines. What types of additional data and findings will be needed before researchers are willing to offer a more balanced and nuanced assessment of the consequences of HT? Will it be feasible for revised guidelines to comprise recommendations that depend on a woman’s age, duration since menopause, duration of use, or degree of vascular disease? This concern is especially important given that almost half of gynaecologists did not believe the initial WHI results, many were confused about the findings, and vast numbers of women did not have accurate information about HT-associated risks.

Have we checkmated ourselves?

Paradoxically, large-scale trials needed to address these issues are unlikely to be forthcoming. In a prescient commentary examining the differences in findings between RCTs and observational studies of HT, Michels asks: “Have we checkmated ourselves?” Now, 6 years later, it seems that we have. Implementation of trials involving random assignment of relatively young menopausal women to HT regimens would be hindered by ethical issues associated with the premature termination of earlier trials, by huge costs and complexities (e.g. the study would need to be large or of long duration to obtain sufficient numbers of clinical events in younger women) and by the biases discussed above that plagued the WHI and other RCTs. At the same time, there is a pressing need for research establishing the biological pathways linking HT to cardiovascular disease, including the role of the timing of HT initiation and duration of use. The WHI-CACS is a significant step in the right direction and several small trials (e.g. the Kronos Early Estrogen Prevention Study (KEEPs) and the Early versus Late Intervention Trial with Estrogen (ELITE)) are underway. Nevertheless, in the light of termination of HT use in the WHI, which includes the ancillary WHI-CACS study, and major obstacles to the development of future large-scale trials, observational studies will need to provide many of the answers.

While once thought to be the trigger for the demise of observational epidemiology, the ongoing HT controversy may accomplish the reverse. Researchers have come to the defence of observational studies in the past, supporting their value for generating new hypotheses, evaluating long-term complex treatment regimens, and providing knowledge in the many circumstances in which RCTs are not ethically or logistically possible. The HT saga is reminding us that observational studies and RCTs provide checks on one another. In the absence of high-quality cohort studies, it would have been difficult, if not impossible, to question the WHI conclusion that HT leads to more negative than positive health outcomes and higher coronary risk for the typical user. Moreover, without comparative studies derived from both observational and experimental data, it would have been challenging to identify how the health risks of HT vary by timing and duration. High-quality observational studies may now be our best hope for gaining additional insights into the biological mechanisms underlying the apparent changes in the effects of HT over a postmenopausal women’s life course.

Acknowledgements

The author would like to thank Ronald Brookmeyer, Dana Glei, Graham Lord and James Trussell for their helpful advice and Diana Sacké for assistance with the manuscript.

Statements on funding and competing interests

Funding Partial support came from the Demography and Epidemiology Unit of the Behavioral and Social Research Program of the National Institute on Aging (RO1AG16790) and the Eunice Kennedy Shriver National Institute of Child Health and Human Development (R24HD047879).

Competing interests None identified.

References

BOOK REVIEW


This book has been published on behalf of the College of William and Mary and the Colonial Williamsburg Foundation in the USA, both of which coincidentally I came across this summer while in Virginia. It is very clear that Americans have a deep relationship with their colonial heritage and take huge pride in authenticating and recreating their origins, as anyone who has visited Williamsburg or small town historic sites has witnessed. Similarly, this book has been meticulously researched from a wealth of sources and challenges old conceptions on fertility and family planning in colonial times. The author’s central thesis is that American women began to reject the lifetime of childbirth and started to limit births more than 100 years in advance of Western Europe, with the exception of France; an alternative ‘American revolution’ that was invented and implemented by women themselves.

This book is not easy bedtime reading material and has numerous footnotes with historical references. It relates to many diverse ethnic, cultural and religious groups such as Quakers, the Pennsylvania Dutch, rural and urban populations. There are attractive illustrations of women and family groups, often portrayed with fruit and flowers as iconic props representing ‘female promise and procreation’. There is a complex association throughout with the demographics of enslaved women; I would have liked to see more chapters specially analyzing their lives and patterns of childbirth. Thomas Jefferson, the third USA president, took up with a slave, Sally Hemmings, after the death of his wife in 1782. Modern DNA testing has suggested his lineage continues in the current African-American population although, as a consummate politician, he did not commit any carnal relationship with a slave at the time.

The birth control chapter states that most historians conclude that contraception was not significant in colonial America. The author disputes this and found examples to the contrary, but mainly referencing herbal remedies from diaries, chemists and doctors’ notes. She found little reference to sexual abstinence other than with nursing mothers. Given the lack of effective contraception, it is not surprising that the revolution limited family size.

In summary, if you are writing a PhD on historical patterns of fertility in North America this book will be a godsend. For the rest of us, it may come with a little dusty on the bookshelf but was an interesting read over the Christmas break nevertheless.

Reviewed by Alsia Gebbie, FRCOG, FFSRH, Consultant in Community Gynaecology, NHS Lothian, UK
New evidence rekindles the hormone therapy debate

Noreen Goldman

*J Fam Plann Reprod Health Care* 2010 36: 61-64
doi: 10.1783/147118910791069277

Updated information and services can be found at:
http://jfprhc.bmj.com/content/36/2/61.citation

**Email alerting service**

Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

**Notes**