

# What Can We Learn from Design Faults in the Women's Health Initiative Randomized Clinical Trial?

Orkun Tan, M.D., S. Mitchell Harman, M.D., Ph.D., and Frederick Naftolin, M.D., Ph.D.

## Abstract

*Design faults resulted in the inability of the Women's Health Initiative (WHI) randomized clinical trial to test the level of cardioprotection conferred by timely hormone treatment of women seeking help for menopausal complaints. Adopting a design constructed around the avoidance of symptomatic subjects and recruitment of older subjects who were more likely to manifest cardiovascular events during the life of the WHI resulted in recruitment of older, sicker subjects than are normally treated for complaints around the time of menopause. The lack of cardioprotection in subjects that began treatment a decade or more after menopause diluted cardioprotection in subjects starting treatment close to the menopausal transition. As a result, despite having the largest number of subjects ever, there were not enough women in the WHI who were comparable to those in the observational trials that showed cardioprotection. This led the WHI to report that there was no cardioprotection in the trial, a position that has been qualified after further analysis.*

*Misapprehension of the initial WHI conclusions by the media, professionals, and regulatory agencies led to a major shift away from menopausal hormone treatment. This remains problematic since the evidence continues to favor cardioprotection and other benefits that are denied under present regulations and guidelines. Regulatory agencies and professional organizations need to better understand the flaws in the WHI design and results in order to properly*

*consider its results and the sustainability of their earlier conclusions and recommendations. Additionally, new trials are needed to test the validity of menopausal hormone-related cardioprotection.*

The WHI is a study that was designed to allow randomized controlled evaluation of three distinct interventions: 1. a low fat eating pattern, hypothesized to prevent breast cancer and colorectal cancer and, secondarily, coronary heart disease; 2) hormone replacement therapy (HRT), hypothesized to reduce the risk of coronary heart disease and, secondarily, to reduce the risk of hip and other fractures, with increased breast cancer risk as possible adverse outcome; and 3. calcium and vitamin D supplementation, hypothesized to prevent hip fractures and, secondarily, other fractures and colorectal cancer.<sup>1</sup> Both estrogen in combination with progesterone (E+P) and estrogen-only (E) arms were terminated prematurely after 5 and 8 years, respectively, although an observational study continues. Briefly, the absolute increase in events for the E+P arm of WHI was as follows: seven more cases of coronary artery disease and eight more cases of invasive breast cancer per 10,000 women per year. However, final results of WHI E+P arm showed that both nominal and adjusted confidence intervals for heart disease or breast cancer either touched or crossed 1 and therefore were not statistically significant. As for the results of the E-only arm, there were five fewer coronary events and seven fewer invasive breast cancer events per 10,000 women per year. The smaller number of coronary and breast cancer events in the E-only arm did not reach a statistical significance.<sup>2</sup> Both arms showed an increased rate of thromboembolic events and stroke. Both arms showed protection against fractures, but with protection against colon cancer only in the E+P arm.<sup>3</sup> These results have been widely generalized as a negative risk-benefit ratio for HRT in menopausal women. The WHI results are at odds with

Orkun Tan, M.D., and Frederick Naftolin, M.D., Ph.D., are from the Department of Obstetrics and Gynecology, New York University School of Medicine, NYU Langone Medical Center, New York, New York. S. Mitchell Harman, M.D., Ph.D., is from the Kronos Longevity Research Institute, Phoenix, Arizona.

*Correspondence:* Frederick Naftolin, M.D., Department of Obstetrics and Gynecology, New York University Langone Medical Center, New York, New York 10016; frederick.naftolin@nyumc.org.

the results of numerous large observational studies that, on average, showed significant (approximately 40% reduction) protection against cardiovascular disease.

### **The main design flaw was to study a population that did not approximate the populations of the observational studies that inspired the WHI**

Why did large observational studies such as the Nurses' Health Study (NHS) show a decrease in coronary heart disease and an increase in breast cancer risk while these findings were not substantiated by WHI? The main differences between WHI and the observational studies that inspired it are the chronological age of subsets, menopausal age (years since the last menstrual period), and the physical condition of the subjects. One of the critical design faults of the WHI is to have accepted to study a different population than in the observational trials in order to avoid dropouts and to have sufficient power to evaluate clinical events rather than progress of disease. For example, the mean age in NHS was 57 years, whereas in WHI study, it was 63 years in the E+P subgroup. Also, the WHI subjects started the hormone therapy (HT) at an average of 12 years postmenopause, in contrast with women in the NHS, who commenced hormones in the perimenopausal and early postmenopausal periods (average age at initiation, 51 years); the latter being consistent with both primary prevention goals and with typical clinical practice. Also, WHI studies combined a small (less than 20% of the study population) healthy group of patients in their early 50s at the start of the study with a much larger study group of patients in their late 50s to late 70s, many of whom can be assumed to have had advanced subclinical disease; in the E+P arm of WHI study, only 33% of hormone-treated and control subjects were 50 to 59 years old, and only 16% to 17% were within 5 years of menopause at the time of enrollment.<sup>4</sup>

### **Purposely avoiding symptomatic subjects furnished a non-random group from which an older population was selected**

Another study design fault is that the WHI study avoided enrolling subjects with symptoms (vasomotor episodes) that would betray the placebo and might increase the rate of dropouts. In observational studies, subjects are self-selected by symptoms and then stratified to those who received HRT or nothing/placebo. In WHI, the trial assigned treatment irrespective of symptoms in addition to selecting women who were 12 or more years, on average, postmenopausal.

### **Systematic absence of events in the placebo group in year five artificially inflated the Hazard ratio at a time when the drug group may have been having fewer events**

The placebo results in year five were across-the-board, approximately half of the numbers of coronary events observed

in year four and six. The latter caused serious problems, as it increased the Hazard ratio falsely and triggered action by the drug safety monitoring board.<sup>3</sup> This apparent loss of a large number of events has not yet been explained, despite its impact on the Hazard ratios.

### **Biological plausibility was not achieved**

WHI also accepted the biological implausibility that the development of heart disease would have been unaffected by age. Clinical cardiovascular disease has a long latency period, and atheroma formation and endothelial dysfunction precede clinical cardiovascular events by many years. Early initiation of estrogen replacement has been shown to inhibit atherosclerosis and the response to vascular injury in a series of animal models.<sup>5,6</sup> This is not surprising in light of the presence of both estrogen receptors and estrogen synthetase in human coronary vessels.<sup>7</sup> However, recent studies suggest that this benefit is lost if initiation of estrogen replacement is delayed until years after menopause. This may reflect estrogen's ability to prevent lesion formation but inability to prevent coronary thrombosis and occlusion in the presence of already-established lesions.<sup>7</sup> There is a clear relationship between absolute calcium scores and severity of coronary artery disease. As shown by Raggi and colleagues,<sup>8</sup> the atherosclerotic plaque burden measured by coronary calcium in asymptomatic women undergoing electron beam tomography revealed increased plaque burden in patients between 60 to 70 compared to patients between 45 to 54 years old.

### **In order to study the outcome of disease it was necessary to study an older, less healthy population**

Studying the outcomes rather than the progress of the disease in the WHI furnished an inappropriate subject population that did not settle the issue of whether E or E+P, as used in the observational studies, is cardioprotective. It should not be surprising that salutary effects of estrogen on cardiovascular disease may require early administration and a long observation period before the better-maintained cardiovascular health of the treated women becomes apparent. These observations suggest that the appropriate study group for postmenopausal cardioprotection is newly menopausal women who receive estrogen for some years, as was the case in the observational studies.<sup>9,10</sup>

### **Not studying the correct population may impact the ability to have enough subjects or effects for statistical power**

Randomized controlled trials are very powerful investigative tools that are limited in their interpretation to populations studied in the randomized controlled trial. Therefore, to assess the power of the WHI trial to resolve such questions, it is necessary to know the number of subjects being observed, the homogeneity of each trial group, and whether there should be subgrouping analysis because of skewed

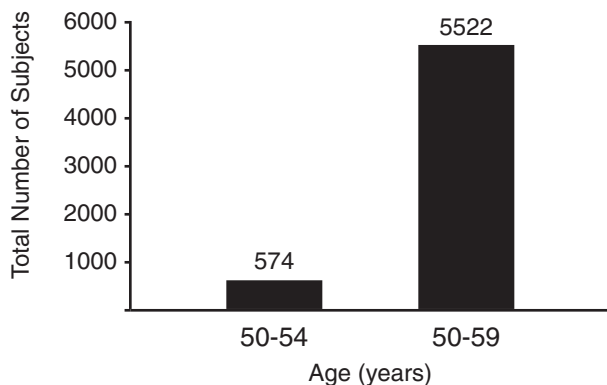
distribution of subjects that could obscure age-related occurrences of cardiovascular events within the larger group of subjects. That is, did WHI data have enough power to test the cardioprotective effects of HRT in women in the menopausal transition (age 49 to 55)? The answer is, no. In fact the WHI study was approximately ten-fold underpowered to test the cardioprotective effects of HRT in women in the menopausal transition.<sup>4</sup>

**What was the number of oppositional WHI subjects vis-a-vis the observational studies?**

Perhaps of greatest importance is the information that, by design, although there were only approximately 2000 moderate to severely symptomatic women in the aggregate E+P and placebo subjects, only a total of 574 women in both groups were 50 to 54 years old and moderately to severely symptomatic<sup>11</sup> (Fig. 1). The power analysis for 50 to 54 years old WHI subjects with moderate to severely symptomatic group had 287 subjects per group.<sup>4</sup>

**Inadequate power to detect anticipated effects in the appropriate population of subjects**

Detecting differences in the occurrence of infrequent events is problematic in small sample sizes, such as those present in the 50- to 54-year-old symptomatic women in the WHI. For example, age-specific data from the NHS indicates that the incidence of cardiac events in the 50- to 54-year-old population is 53/100,000 per year.<sup>12</sup> This translates to 0.73 expected events in 275 women over a 5-year period. Even if there was a several fold difference in the number of events between the E+P and placebo groups, the small sample size would make it very unlikely that a statistically significant difference could be detected: a power analysis indicates that



**Figure 1** Total number of 50- to 54-year-old moderate to severely symptomatic subjects in the E+P and placebo groups compared with the total number of subjects in the 50- to 59-year-old groups. (Data from Naftolin F, Taylor HS, Karas R, et al. The Women’s Health Initiative could not have detected cardioprotective effects of starting hormone therapy during the menopausal transition. *Fertil Steril.* 2004 Jun;81(6):1498-501.)

assuming 0 events in the placebo group and twice the number of expected events in the E+P group, it would require greater than 4000 women in each arm of the study to detect such a difference with statistical significance. Moreover, given that there was a 42% dropout rate, as reported by the WHI,<sup>3</sup> the number of subjects needed per group rises to almost 9000; the WHI had only 287 per group. Stated another way, using the number of symptomatic, newly menopausal women present in the WHI, it would require at least a nine-fold increase in the number of events in the trial arm to achieve statistical significance. The excess events for the entire trial, including women 55 to 80 years of age, was less than one-fold.<sup>13</sup>

Thus, the WHI was more than 10-fold underpowered to detect a change in clinical cardiovascular events in the patient population most likely to be capable of receiving benefit. With 287 (574/2) subjects per group, the WHI could not reasonably be expected to provide useful information regarding the cardioprotective effects of E+P in moderately to severely symptomatic women who were 50 to 54 years old at the start of the trial. In support of this interpretation, Manson and associates<sup>13</sup> reported a non-statistically significant decreased relative risk of cardiovascular events in hormone therapy users who were less than 10 years from the onset of menopause. Had the study been sufficiently powered, this decreased relative risk might have achieved statistical significance. Had the investigators segregated those closer to the menopausal transition or had the study included sufficient numbers of newly menopausal women, they might well have observed a further decrease in relative risk.

**Non-random assignment of subgroups within the entire group makes that a non-randomized population**

Another critical design fault by WHI was studying a selected subset group from the larger recruitment to determine if E+P is associated with an increased incidence of dementia and mild cognitive impairment in postmenopausal women. The investigators concluded that, overall, 61 women were diagnosed with “probable dementia,” 40 (66%) in the estrogen plus progestin group compared with 21 (34%) in the placebo group. The Hazard ratio for probable dementia was 2.05.<sup>14</sup> In addition to the non-randomness of the selection, no neurologic examination was performed at the initial enrollment, which further clouds the issues.

**Conclusion**

In conclusion, the WHI failed to meet its (never clearly stated) objective to test for the cardioprotective effect of hormonal treatment in menopausal subjects equivalent to those in the observational trials. It also did not resolve the question of quality of life improvement, protection against dementia, or the relationship of estrogen treatment to breast cancer prevalence in the context of the women in the observational studies that inspired the WHI. It did show for the first time that menopausal hormone treatment decreases

the incidence of non-vertebral (femoral neck) fractures in aging women. The latter is a subject beyond the scope of this publication, but one that has been overlooked in the misapprehension of the results of the WHI.<sup>15</sup>

It is critical that women and caregivers understand that the WHI study was not aimed at and was not powered to examine women at the menopausal transition or in early menopause, and its results therefore are not directly applicable to the usual population of women who seek consultation at the menopause clinics.<sup>16</sup> This has been the thrust of several important publications from members of the WHI investigator group and professional societies.<sup>13,15,17</sup>

In the absence of an adequately powered study of women in the menopausal transition, it is not appropriate to define either clinical management of symptomatic 50- to 54-year-old women or to mandate discontinuation of appropriately initiated hormone therapy on the basis of the available data from the WHI. Since many observational trials have already indicated a cardioprotective effect of early estrogen treatment, well-designed prospective randomized controlled trials should provide better understanding of the risks and benefits of hormone replacement therapy in peri- and postmenopausal women.<sup>18</sup>

#### *Disclosure Statement*

None of the authors have a financial or proprietary interest in the subject matter or materials discussed, including, but not limited to, employment, consultancies, stock ownership, honoraria, and paid expert testimony.

#### **References**

1. Women's Health Initiative Study Group. Design of the Women's Health Initiative Clinical trial and observational study. *Control Clin Trials*. 1998;19:61-109.
2. The Women's Health Initiative Steering Committee, effects of conjugated equine estrogen in postmenopausal women with hysterectomy. *JAMA*. 2004;291:1701-12.
3. 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;288:321-33.
4. Naftolin F, Taylor HS, Karas R, et al. The Women's Health Initiative could not have detected cardioprotective effects of starting hormone therapy during the menopausal transition. *Fertil Steril*. 2004 Jun;81(6):1498-501.
5. Mikkola S, Clarkson TB. Estrogen replacement therapy, atherosclerosis, and vascular function. *Cardiovasc Res*. 2002;53:605-19.
6. Haynes MP, Li L, Russell KS, Bender JR. Rapid vascular cell responses to estrogen and membrane receptors. *Vasc Pharmacol*. 2002;38:99-108.
7. Diano S, Horvath TL, Mor G, et al. Aromatase and estrogen receptor immunoreactivity in the coronary arteries of monkeys and humans. *Menopause*. 1999;6:21-8.
8. Raggi P, Khan A, Arepali C, Stillman AE. Coronary artery calcium scoring in the age of CT angiography: what is its role? *Curr Atheroscler Rep*. 2008 Oct;10(5):438-43.
9. 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;325:756-62.
10. Langer RD. Legend meets reality: estrogen plus progestin and coronary heart disease in the Women's Health Initiative. *Menopausal Med*. 2003;10:5-7.
11. Hays J, Ockene JK, Brunner RL, et al. Effects of estrogen plus progestin on health-related quality of life. *N Engl J Med*. 2003;348:1839-54.
12. 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 Sep 12;325(11):756-62.
13. Manson JE, Hsia J, Johnson KC, et al. Estrogen plus progestin and the risk of coronary heart disease. *N Engl J Med*. 2003;349:523-34.
14. Shumaker SA, Legault C, Rapp SR, et al. Estrogen plus progestin and the incidence of dementia and mild cognitive impairment in postmenopausal women: the Women's Health Initiative Memory Study: a randomized controlled trial. *JAMA*. 2003 May 28;289(20):2651-62.
15. Utian WH, Archer DF, Bachmann GA, et al. Estrogen and progestogen use in postmenopausal women: July 2008 position statement of The North American Menopause Society. *Menopause*. 2008;15:584-602.
16. Naftolin F, Schneider HP, Sturdee DW. Executive Committee of the International Menopause Society. Guidelines for the hormone treatment of women in the menopausal transition and beyond. *Climacteric*. 2004;7:8-11.
17. 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*. 2006;15:35-44.
18. Brinton EA, Hodis HN, Merriam GR, et al. Can menopausal hormone therapy prevent coronary heart diseases? *Trends Endocrinol Metab*. 2008;19:206-12.