In 2002, the Women’s Health Initiative (WHI) clinical trial reported that combined estrogen-plus-progestin hormone therapy did not prevent coronary heart disease in women (1). Combined estrogen-plus-progestin therapy increased the risk of stroke by a factor of 1.4 on average and doubled the risk of venous thromboembolism.

Observational research up to the time of the WHI suggested that the relative risk of coronary heart disease was 0.50–0.65 in hormone users compared with nonusers (2, 3). Based on observational research, the relative risk of stroke was considered less than or near 1.0 for hormone therapy (2). Both observational studies (4) and a randomized trial, the Heart Estrogen/progestin Replacement Study (HERS) (5), showed a relative risk of 1–2 for venous thromboembolism in hormone users.

The WHI clinical trial joined a body of experimental evidence showing no effect of postmenopausal hormone therapy on coronary heart disease clinical endpoints or measures of subclinical coronary atherosclerosis (6–16). In 1998, the Heart Estrogen/progestin Replacement Study reported no beneficial effect of hormone therapy on morbidity or mortality from coronary heart disease in women with established coronary disease (6). Seven other randomized trials of the effect of hormone therapy on arterial disease endpoints published in 2000–2002 found no effect of hormone therapy (7–13). A 1997 meta-analysis of small randomized trials of hormone replacement found that hormone therapy did not prevent cardiovascular disease (14). Follow-up of participants in the Heart Estrogen/progestin Replacement Study (15) showed no benefit of hormone therapy for any cardiovascular endpoint. The estrogen-only arm of the WHI clinical trial was terminated early after showing that hormones did not prevent coronary heart disease (16).

In this issue, Prentice et al. (17) report an analysis of data from the WHI clinical trial and the WHI observational study. The WHI clinical trial and observational study were conducted by the same group of researchers, in the same time frame, and had similar procedures for ascertaining and monitoring events. Prentice et al. explore the reasons for the differences in estimates of coronary heart disease, stroke, and venous thromboembolism risk for hormones between the WHI clinical trial and the observational study. They attempt to resolve the discrepancies by statistical adjustment.

TRADITIONAL CONFOUNDERS

Table 1 below shows how adjustments of the WHI clinical trial data and observational study data affect hazard ratios for coronary heart disease, stroke, and venous thromboembolism. Given the differences in age between hormone users and nonusers in the WHI observational study, it is not surprising that age adjustment has a large effect on hazard ratios for all three endpoints. Adjustment for age is standard in observational research. After adjustment for other confounders, including known cardiovascular risk factors and education, the hazard ratios for coronary heart disease, stroke, and venous thromboembolism for the WHI observational study all move closer to those from the WHI clinical trial.

Prentice et al. (17) point out that education was related strongly to coronary heart disease risk in the WHI. Education is a surrogate for socioeconomic status. Inclusion of a measure of socioeconomic status as a potential confounder in the analysis of observational data for virtually any exposure/disease relations seems fundamental. Humphrey et al. (18) reported that only three of 10 observational studies

<table>
<thead>
<tr>
<th>Adjustment</th>
<th>Coronary heart disease</th>
<th>Stroke</th>
<th>Venous thromboembolism</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Clinical trial</td>
<td>Observational study</td>
<td>Clinical trial</td>
</tr>
<tr>
<td>None</td>
<td>1.18†</td>
<td>0.50†</td>
<td>0.42†</td>
</tr>
<tr>
<td>Age only</td>
<td>1.21†</td>
<td>0.71†</td>
<td>0.61§</td>
</tr>
<tr>
<td>Age and traditional confounders</td>
<td>1.27¶</td>
<td>0.87¶</td>
<td>0.70§</td>
</tr>
<tr>
<td>Age and traditional confounders by time since estrogen-plus-progestin hormone therapy initiation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;2 years</td>
<td>1.68#</td>
<td>1.12#</td>
<td></td>
</tr>
<tr>
<td>2–5 years</td>
<td>1.25#</td>
<td>1.05#</td>
<td></td>
</tr>
<tr>
<td>&gt;5 years</td>
<td>0.66#</td>
<td>0.83#</td>
<td></td>
</tr>
<tr>
<td>Summary</td>
<td>0.93§</td>
<td>0.76§</td>
<td>0.84§</td>
</tr>
</tbody>
</table>

* Annualized incidence ratios or hazard ratios.
† Source: table 1, Prentice et al. (17).
‡ Calculated as the ratio of annualized incidence ratios using the data from table 1, Prentice et al. (17).
§ Source: text citation, Prentice et al. (17).
¶ Source: table 4, Prentice et al. (17).
# Source: table 5, Prentice et al. (17).

examining coronary heart disease risk in relation to current hormone therapy adjusted for a measure of socioeconomic status. A summary estimate of the relative risk of coronary heart disease in current hormone users based only on studies that adjusted for a measure of socioeconomic status was 0.97 (95 percent confidence interval: 0.82, 1.16); when based on studies that did not adjust for a measure of socioeconomic status, the relative risk estimate was 0.71 (95 percent confidence interval: 0.64, 0.78). Reports on coronary heart disease risk from the most influential observational study of hormone therapy—the Nurses’ Health Study—did not adjust for socioeconomic status (19–25).

It has been argued that adjustment for socioeconomic status is not necessary in studies that enroll subjects who are homogeneous (26). Routine presentation of results adjusted for socioeconomic status would bolster this kind of argument. However, the argument raises important questions. Why did it take so long to recognize that many studies failed to adjust for socioeconomic status? Why were studies that adjusted for socioeconomic status disregarded by most observers?

TIME COURSE OF EXPOSURE

Prentice et al. (17) developed a model that adjusts for traditional covariates and takes into account the time course of exposure, including the time since initiation of use in the most recent episode of use. The model brings the WHI observational data and the WHI clinical trial data into statistical agreement for two of the three endpoints considered, namely, coronary heart disease and venous thromboembolism. Although the hazard ratios are in statistical agreement, some point estimates are substantively different (table 1). Prentice et al. continued to try different adjustments, because the WHI clinical trial results showed that the WHI observational study results were still incorrect after the initial set of adjustments was complete. Without the clinical trial, it would not be clear when to stop adjusting or whether prior adjustments were successful. As stated by Prentice et al., after adjustment for traditional confounders (including education as a measure of socioeconomic status), the WHI clinical trial found a hazard ratio for stroke of 1.21, whereas the observational study found a hazard ratio of 0.86 for stroke. For stroke, Prentice et al. state that the hazard ratios in the observational study compared with the clinical trial differed significantly (p < 0.01) even after further adjustment for the time course of estrogen-plus-progestin hormone therapy.

FURTHER INFORMATION ON STROKE AND CONCERN ABOUT DEMENTIA

Both a nonsystematic review and a recent systematic review and meta-analysis of clinical trial data concluded that hormone therapy increases the risk of stroke (27, 28). Stroke is overlooked as a major cause of morbidity and mortality in women, and its frequency relative to coronary heart disease is underappreciated. The relative contributions of coronary heart disease and stroke to arterial disease in the WHI—roughly equal—are typical for women aged 55–70 years living in the United States now.

In contrast with a number of observational studies (29–36), which found substantial reductions in the risk of dementia in relation to hormone therapy, the Women's Health Initiative Memory Study (WHIMS) found that hormone therapy increased the risk of dementia by a factor of about two (37, 38). This effect may be related to subclinical brain infarction (37–39). The long-term effects of hormones on the risk of dementia remain unknown.

CONCLUSION

The analysis by Prentice et al. (17) confirms that there is no overall benefit of hormone therapy in preventing cardiovascular disease. Randomized trials of hormone therapy changed the practice of medicine. Hormone therapy should not be used to prevent cardiovascular disease. The conclusion that hormone therapy prevents cardiovascular disease, which was based on observational research, was wrong. Possible explanations for the error include the apparent consistency of the cohort studies and a belief that the estimated size of the effect was too large to be explained by confounding.

Most of what we know about causation from John Snow to the present comes from observational studies. Observational studies will remain a staple in the field. Prentice et al. remind us to pay more attention to fundamentals in observational research. They highlight some ways that observational studies could be better designed and better analyzed. However, observational studies are not a substitute for clinical trials no matter how sophisticated the statistical adjustments may seem.

ACKNOWLEDGMENTS

Conflict of interest: none declared.

REFERENCES