In this week's reading paper, the authors attempted to explain the discrepancies between the Women's Health Initiative (WHI) and the Nurses' Health Study (NHS). The WHI randomized trial found increased coronary heart disease (CHD) risks in postmenopausal hormone users whereas the NHS observational study found reduced CHD risks in the hormone users. They reconceptualized the NHS study as a sequence of trials and estimated the intention-to-treat effects between the two studies. Based on the results, the difference between the NHS and the WHI study could be largely explained by unmeasured confounders.

The main takeaway of this work is that it is hard to eliminate the unmeasured confounding effect without randomization. Aside from technical highlights, what impresses me most is that the authors are able to reuse data from the NHS trials which recruited a large cohort size from decades of years ago. I have no idea what process is required for reusing the dataset but I guess it must involve a lot of effort as a considerable amount of PHIs are included in the data. Without the sharing of the NHS data, it is impossible to address the discrepancies between the NHS and the WHI study as in this paper. Accelerating the sharing of biomedical data without leaking PHIs can be beneficial to researchers. I have been using open-access datasets such as MIMIC-III/IV, All of Us and N3C, which I feel is quite helpful. Hopefully, more open-access datasets with larger sample sizes and more modalities will become available in the future.