# Instrumental Variable Methods Reveal Larger Effects of Menopausal Hormone Therapy in the Landmark Women's Health Initiative Clinical Trial

*By* JOSHUA ANGRIST, AMANDA E. KOWALSKI, LJUBICA RISTOVSKA, MARCIA STEFANICK*

* Angrist: Massachusetts Institute of Technology (e-mail: angrist@mit.edu); Kowalski: University of Michigan (e-mail: aekowals@umich.edu); Ristovska: Yale University (e-mail: ljubica.ristovska@yale.edu); Stefanick: Stanford University (e-mail: stefanick@stanford.edu). Acknowledgments: We thank Saumya Chatrath, Zoey Chopra, Tory Do, Bailey Flanigan, Anastasiya Karpova, Pauline Mourot, Srajal Nayak, Dominik Piehlmaier, Sukanya Sravasti, and Matthew Tauzer for research assistance. Marcia Stefanick is the Stanford Principal Investigator of the WHI. This study was deemed exempt by the NBER IRB. It has been approved by the WHI P&P Committee. The WHI program is funded by the National Heart, Lung and Blood Institute, NIH through 75N92021D00001, 75N92021D00002, 75N92021D00003, 75N92021D00004, 75N92021D00005. The authors thank the WHI investigators and staff for their dedication, and the study participants for making this research possible. A list of WHI investigators can be found at https://www.whi.org/doc/WHI-Investigator-Short-List.pdf.

Menopausal hormone therapy (MHT) was widely prescribed in the U.S. during the 1990s to alleviate symptoms of menopause. Observational studies suggested that MHT might protect against coronary heart disease, prompting debate over whether MHT should be used more widely for preventive purposes among women without symptoms. To assess the causal impact of MHT on healthy postmenopausal women, two Women's Health Initiative (WHI) clinical trials randomized women with and without a uterus to MHT or placebo.

The trial for women with a uterus was stopped early at the recommendation of the data safety and monitoring board because a global index of health outcomes reached a critical statistically significant threshold, as did a test statistic for effects on invasive breast cancer. A landmark publication showed that women assigned MHT faced higher risks of coronary heart disease, invasive breast cancer, stroke, and pulmonary embolism (Writing Group for the Women's Health Initiative Investigators, 2002). The results were widely publicized. Although 39% of U.S. women aged 52 to 65 used MHT in 2001, usage declined precipitously to 8% by 2009 (Yang and Toriola, 2024). New breast cancer cases, particularly estrogen-receptor-positive cases, also decreased sharply in aggregate U.S. statistics (Ravdin et al., 2007).

Subsequent findings using longer follow-up data show that assignment to MHT reduced the risk of colorectal cancer and hip fractures (Manson et al., 2013), reigniting controversy. A 2024 review from the WHI attempts to reconcile these contradictory findings, recommending against the use of MHT for chronic disease prevention while at the same time stating that "menopausal hormone therapy is appropriate to treat bothersome vasomotor symptoms among women in early menopause, without contraindications, who are interested in

taking hormone therapy" (Manson et al. 2024). A 2024 *New York Times* article argues that findings from the WHI "are now considered to have been overblown" and that the increase in breast cancer was "very small" (Friedman, 2024).

We revisit the impact of MHT, applying instrumental variable (IV) methods to data from the WHI. Published estimates focus on intention-to-treat (ITT) estimates comparing outcomes based on random assignment. ITT methods preserve the balance targeted by random assignment; however, when treatment taken deviates from the assignment, ITT understates the causal effects on treated trial participants. Many WHI trial participants deviated from their assignment. Almost 40% of women assigned to MHT discontinued study medication before the trial was stopped, and 9% of women randomized to placebo initiated MHT. ITT estimates ignore such nonadherence and crossovers.

IV methods instead capture causal effects of MHT use for trial participants induced to take MHT by the trial. IV accounts for both nonadherence (failure to use MHT in the trial arm assigned MHT) and crossovers (MHT use in the group assigned placebo). Applications of IV to clinical trials include studies of screening for colorectal cancer (Angrist and Hull, 2023) and breast cancer (Kowalski, 2023). When treatment can be coded as a dummy variable distinguishing participants by treatment status, IV estimates necessarily exceed the corresponding ITT estimates. IV methods also accommodate treatment coded in terms of years of exposure to (i.e., use of) MHT. In this case, IV delivers the average causal effects of an additional year of treatment.

## I. The Women's Health Initiative

The WHI is a long-term national health initiative studying strategies to prevent major diseases among postmenopausal women. WHI includes both observational studies and large-scale clinical trials. The Estrogen plus Progestin (E + P) trial, the focus of this paper, studies the effects of CEE combined with medroxyprogesterone acetate (MPA) in women with an intact uterus at baseline.[1]

The E + P trial enrolled 16,608 women from 1993-98, with 8,506 randomized to MHT and 8,102 randomized to placebo. As expected, baseline health and demographic characteristics are similar in the group assigned

[1] The Estrogen-only (E-alone) trial studied the effects of conjugated equine estrogens (CEE) on women with a prior hysterectomy and also ended prematurely. Long-term results from the E-alone trial show statistically significant *reductions* in breast cancer mortality (Manson et al. 2017).

MHT and the group assigned placebo.[2] In July 2002, the trial was stopped and participants were advised to discontinue study medication.

## II. Intent to Treat (ITT) vs. Instrumental Variables (IV) Methods

Two groups of trial participants are treated. One group – represented by participants assigned control – take MHT whether or not they are assigned to receive it. Another – called *compliers* – take MHT when assigned the intervention but not otherwise. As a theoretical matter, IV methods capture causal effects on compliers (Imbens and Angrist 1994, Angrist, Imbens, Rubin 1996)[3]. We implement IV methods using random assignment as an instrument for MHT use.

IV estimates are constructed as a ratio. The denominator of this ratio, called the *first stage*, is given by the difference in MHT use between those assigned MHT and those assigned placebo. This is an ITT estimate of the effects of random assignment to the intervention on treatment received. The numerator, called the *reduced form* in an IV framework, is the ITT effect for a given outcome. IV divides ITT estimates for downstream outcomes by first-stage estimates. Intuitively, IV attributes changes in outcomes induced by random assignment to higher treatment rates among those assigned treatment rather than placebo.

IV analysis includes everyone initially randomized, regardless of adherence or crossover behavior. The landmark WHI study reports results that censor outcomes six months after participants become non-adherent, defined as less than 80% adherence to study medication or stopping study medication completely (Writing Group for the Women's Health Initiative Investigators, 2002). This is potentially misleading since those who are censored may have latent health systematically different from those retained. In fact, any analysis that conditions on treatment received, such as widely reported per-protocol estimates of randomized trials like WHI, are almost certainly biased. As-treated estimates effectively discard random assignment. IV includes data on all participants and, like ITT, relies on comparisons by random assignment in both the first stage and reduced form.

[2] Participant characteristics at baseline include demographics, reproductive and medical history, history of hormone use, health behaviors, and health measures.

[3] The landmark publication uses the term "compliers" to refer to those who adhere to their assigned medication without invoking a counterfactual (Writing Group for the Women's Health Initiative Investigators, 2002). For clarity, we refer to those women as "non-adherent."

## III. Data and Specifications

Our analysis begins with treatment coded as a dummy variable indicating whether a participant took MHT for at least one year.[4] We also construct an ordered treatment variable measuring the number of years a participant took MHT. These measures of MHT use extend up to 11 years after randomization (2005 at the latest) for the participants that enrolled the earliest. All specifications control for age groups at randomization, trial participation, and randomization arm in other WHI clinical trials.[5]

## IV. Results

Table 1 reports results by the end of the intervention phase on July 7, 2002. Column (1) in the top panel shows that participants randomized to MHT were 82 percentage points more likely to take MHT during the intervention phase. This estimate implies that IV estimates are 22% (=100/82 - 1) larger than corresponding ITT estimates for each downstream outcome. In other words, considering the impact of being treated with MHT and not just the intent to treat with MHT increases the estimated magnitudes of all risks and benefits by 22%. The second row in the first panel reports that participants randomized to MHT had an average of 3.7 additional years of MHT use relative to participants assigned placebo. Median follow-up was 5.6 years.

To shed light on the absolute magnitudes of risks and benefits with IV vs. ITT, the bottom panel replicates results from Manson et al. (2013), expressed as ITT estimates.[6] In the first row, we combine outcomes into the global index that precipitated the end of the study. Column (1) reports a statistically significant ITT of 119 additional events per ten thousand women assigned to MHT. However, the corresponding IV estimate in column (2) implies an increase of 145 events per ten thousand women treated with MHT. This estimate is 22% larger, as discussed previously.

[4] Variables include data from three sources: adherence to study medication, open-label MHT use within the study, and MHT use through own physician. While data on adherence to study medication and open-label MHT use were collected annually up to 11 years since randomization, data on MHT use through own physician was collected routinely during visits at years 1, 3, 6, and 9 since randomization. Almost all participants have adherence and open-label data and almost 90% of participants have data from their own physician up to and including year 6 after randomization.

We make a few assumptions to harmonize the data. Using adherence data, we consider women in the E + P arm to have taken MHT each year if they took the study medication continuously for the entire year. Using the other two sources, we assume that if a participant took MHT open-label or through their own physician, they did so for the entire year. We assume that participants who do not have data on MHT use through their own physician at the initial visit year did not take MHT through their own physician in that year. For each year where at least one of the three data sources is non-missing, we impute missing values for the other sources using the last observation carried forward. If all three data sources are missing, we assume the participant is lost to follow-up. We consider participants who are lost to follow-up or deceased to have stopped MHT.

[5] Participants in the MHT clinical trials could also participate in the dietary modification trial or the calcium + vitamin D clinical trial. The WHI landmark paper also controls for prior disease separately for each outcome. To make the construction of our IV results more transparent, we do not control for prior disease because doing so would necessitate a different first stage scaling factor for each outcome.

[6] We use updated data with outcome adjudication through 2023.

TABLE 1— ITT VS. IV DURING THE INTERVENTION PHASE

| | ITT | IV: took MHT | IV: years on MHT |
|---|---|---|---|
| | (1) | (2) | (3) |
| *First stage outcomes* | | | |
| Took MHT | 0.820*** | | |
| | (0.004) | | |
| Years on MHT | 3.662*** | | |
| | (0.024) | | |
| *Outcomes per 10,000* | | | |
| Global index | 119*** | 145*** | 32*** |
| | (46) | (56) | (12) |
| Invasive breast cancer | 50** | 61** | 13** |
| | (23) | (27) | (6) |
| Stroke | 51*** | 62*** | 14*** |
| | (20) | (24) | (5) |
| Pulmonary embolism | 52*** | 63*** | 14*** |
| | (13) | (16) | (4) |
| Colorectal cancer | -35*** | -43*** | -9*** |
| | (14) | (17) | (4) |
| Endometrial cancer | -5 | -6 | -1 |
| | (9) | (11) | (2) |
| Hip fracture | -30** | -37** | -8** |
| | (14) | (17) | (4) |

*Notes:* MHT = Menopausal hormone therapy. Standard errors in parentheses. ***p<0.01, **p<0.05, *p<0.10. IV is the ratio of the outcome ITT and first stage ITT. The intervention phase ended on July 7, 2002. The global index includes all outcomes shown plus coronary heart disease and death. All specifications control for age groups at randomization and trial participation and randomization arm in dietary modification and calcium + vitamin D trials. MHT use was defined using study medication adherence, open-label MHT use, and MHT use through own physician.

Risks of MHT captured within the global index increase by the same scaling factor when using IV. The IV estimate in column (2) implies a statistically significant increase of 61 cases of invasive breast cancer per ten thousand women who took MHT during the intervention phase. The increases in stroke and pulmonary embolism are each similar in magnitude to the increase in breast cancer, and they are statistically significant.

Benefits of MHT also increase by the same scaling factor. IV estimates show that treatment with MHT decreases cases of colorectal cancer by 43 per ten thousand and hip fractures by 37 per ten thousand during the intervention phase.

The difference between reported and IV estimates is even larger if assessing risk and benefits per year. The landmark WHI trial reported additional risks of invasive breast cancer of 9 per 10,000 women per year annualized over intervention phase, which underestimates the treatment effect of MHT if women did not adhere to their assignment for the entire follow-up. The IV estimates using treatment coded as number of years on MHT adjust the ITT estimates by the actual number of years on MHT, regardless of the follow-up duration. These estimates, reported in column (3), indicate that an additional year of MHT use increases the risk of invasive breast cancer by 13 per 10,000 women per year.

Women assigned MHT might have changed their behavior after the study ended to mitigate adverse health outcomes from MHT exposure, so we interpret longer-term results by the end of the postintervention extension phase on September 30, 2010 with caution. We report them in Table 2 for comparison to the literature.

As shown in the first panel, first stage estimates through full follow-up are very similar to estimates from the intervention phase because almost all women discontinued MHT immediately when the trial was stopped. The IV estimate using a binary treatment variable in

column (2) of Table 2 indicates that MHT increases the risk of invasive breast cancer by 135 per ten thousand women during the full follow-up, which is 23% (=100/81.2 - 1) larger than the corresponding ITT estimate. Expressed in per-year terms in column (3), invasive breast cancer increases by 30 per ten thousand women for each additional year of MHT use. This IV estimate is over 3 times larger than the annualized estimate reported in the literature. Even though the ITT estimate of the risk of invasive breast cancer more than doubles with longer follow-up, the annualized estimate reported in the literature remains at 9 additional breast cancer cases per ten thousand per year because it annualizes over a longer follow-up period with a median duration of 12.1 years.

TABLE 2— ITT VS. IV DURING FULL FOLLOW-UP

| | ITT | IV: took MHT | IV: years on MHT |
|---|---|---|---|
| | (1) | (2) | (3) |
| *First stage outcomes* | | | |
| Took MHT | 0.812*** | | |
| | (0.005) | | |
| Years on MHT | 3.644*** | | |
| | (0.028) | | |
| *Outcomes per 10,000* | | | |
| Global index | 110 | 135 | 30 |
| | (68) | (83) | (18) |
| Invasive breast cancer | 110*** | 135*** | 30*** |
| | (33) | (40) | (9) |
| Stroke | 54* | 67* | 15* |
| | (31) | (38) | (8) |
| Pulmonary embolism | 45** | 56** | 12** |
| | (21) | (26) | (6) |
| Colorectal cancer | -35* | -43* | -10* |
| | (20) | (25) | (6) |
| Endometrial cancer | -42*** | -52*** | -12*** |
| | (16) | (19) | (4) |
| Hip fracture | -64** | -79** | -17** |
| | (27) | (33) | (7) |

*Notes:* The full follow-up ended on September 30, 2010. For further details, see the notes in Table 1.

The marked increase in breast cancer in the postintervention extension phase does not extend to stroke and pulmonary embolism, and the stroke result loses statistical significance. Reductions in endometrial cancer, which are not statistically significant in the intervention phase, become statistically significant in the postintervention extension, implying 52 fewer cases overall per ten thousand women who took MHT. The global index increases by the same magnitude as invasive breast cancer during full follow-up, but it is imprecise.

## IV. Conclusion

IV estimates reveal that the health risks from taking MHT, particularly invasive breast cancer and pulmonary embolism, are meaningfully larger than corresponding ITT estimates. The magnitudes of the protective benefits of MHT against endometrial cancer and hip fracture increase by the same factors. Our findings underscore the importance of accounting for compliance with random assignment in clinical trial analyses to provide a more targeted view of the true risks and benefits of medical treatments. They also demonstrate that IV estimates are clinically relevant and that MHT guidelines may need reassessment to reflect the risks and benefits to compliers to the randomized intervention.

## REFERENCES


Angrist, Joshua D., and Peter Hull. 2023. "Instrumental Variables Methods Reconcile Intention-to-Screen Effects Across Pragmatic Cancer Screening Trials." *Proceedings of the National Academy of Sciences* 120: e2311556120.

Angrist, Joshua D., Guido W. Imbens, and Donald B. Rubin. 1996. "Identification of Causal Effects Using Instrumental Variables." *Journal of the American Statistical Association* 91 (434): 444–455.

Friedman, Danielle. 2024. "The Menopause Queen's Gambit." *New York Times*, November 2.

Imbens, Guido W., and Joshua D. Angrist. 1994. "Identification and Estimation of Local Average Treatment Effects." *Econometrica* 62 (2): 467–475.

Kowalski, Amanda E. 2023. "Behaviour Within a Clinical Trial and Implications for Mammography Guidelines." *Review of Economic Studies* 90 (1): 432–462.

Manson, JoAnn E., Rowan T. Chlebowski, Marcia L. Stefanick, et al. 2013. "Menopausal Hormone Therapy and Health Outcomes During the Intervention and Extended Poststopping Phases of the Women's Health Initiative Randomized Trials." *JAMA* 310 (13): 1353–1368.

Manson, JoAnn E., Carolyn J. Crandall, Jean E. Rossouw, Rowan T. Chlebowski, Garnet L. Anderson, Marcia L. Stefanick, Alicia K. Aragaki, Jane A. Cauley, Garnet L. Wells, Andrea Z. LaCroix, Cynthia A. Thomson, Marian L. Neuhouser, Linda Van Horn, Charles Kooperberg, Barbara V. Howard, Lesley F. Tinker, Jean Wactawski-Wende, Sally A. Shumaker, and Ross L. Prentice. 2024. "The Women's Health Initiative Randomized Trials and Clinical Practice: A Review." *JAMA* 331 (20): 1748–1760.

Ravdin, Peter M., Kathleen A. Cronin, Nadia Howlader, Christine D. Berg, Rowan Chlebowski, Eric J. Feuer, Brenda K. Edwards, and Donald A. Berry. 2007. "The Decrease in Breast-Cancer Incidence in 2003 in the United States." *New England Journal of Medicine* 356: 1670–1674.

Writing Group for the Women's Health Initiative Investigators. 2002. "Risks and Benefits of Estrogen Plus Progestin in Healthy Postmenopausal Women: Principal Results from the Women's Health Initiative Randomized Controlled Trial." *JAMA* 288: 321–333.

Yang, Lin, and Temitope A. Toriola. 2024. "Menopausal Hormone Therapy Use Among Postmenopausal Women." JAMA Health Forum 5 (9): e24312.