

Population-Scale Reconstruction Reveals a Systemic Physiological Cliff at the Final Menstrual Period

Corresponding Author: Professor Uri Alon

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

This manuscript presents a large-scale cross-sectional analysis of laboratory data to reconstruct physiological trajectories around the final menstrual period using a deconvolution approach. By analyzing more than 300 million laboratory measurements from two national cohorts, the authors report an abrupt, step-like change across multiple biological systems at menopause and suggest that these changes begin years before the final menstrual period and are attenuated among women using hormone replacement therapy.

The research question is important and the dataset is large, significant strengths of the manuscript. However, the manuscript is poorly organized and the language needs clarity. The Results section is poorly structured and frequently includes methodological explanations, analytical rationale, and variable definitions that belong in the Methods section. The Results should be restricted to objective reporting of findings, without restating why analyses were performed, how variables were constructed, or providing interpretative statements.

Similarly, the Discussion is very brief and largely reiterates the results instead of interpreting them. There is little integration with prior literature and limited discussion of biological plausibility. In addition, there is no section addressing the limitations of the study. Given the cross-sectional nature of the data, potential residual confounding, assumptions underlying the deconvolution approach, and uncertainty in the timing of the final menstrual period, a detailed limitations section is essential. The Methods section would also benefit from reorganization. A clearer structure would begin with a description of the datasets used, including references to their source protocols, followed by eligibility criteria, outcome and variable definitions, data processing steps, and finally the statistical analyses. This order would provide a more logical and transparent flow. The conclusion overstates the implications of the findings. Phrases such as “these findings redefine menopause as a quantifiable, systemic transition” imply a level of conceptual redefinition that is not fully supported. A more appropriate wording would be that the findings may contribute to or “further support” a systemic and quantifiable view of the menopausal transition, rather than claiming a definitive redefinition.

The authors should clarify how they handled women using hormonal medications, such as oral contraceptives, who may not menstruate but are not menopausal. In addition, if women under 30 years who did not respond to questions on regular menstruation were classified as non-menopausal, the choice of this age threshold should be justified, either by rationale or by citation of prior literature.

Since NHANES provides a direct definition of menopause whereas Clalit relies on inferred menopausal timing, the credibility of the results would be strengthened by a sensitivity analysis restricted to NHANES participants with directly defined menopausal status.

The current title does not specify the study design. From the title alone, it is unclear what type of study was conducted.

In the Introduction and Discussion, several statistics and factual statements are cited from review articles rather than from the original studies that first reported these findings. For example, the statement “more than 1.1 billion women are living beyond their final menstrual period” is referenced through secondary sources.

Finally, abbreviations are not used consistently. Some terms are abbreviated before being written in full, and in other cases the full term and abbreviation are reintroduced after already being defined. Consistent use of abbreviations throughout the manuscript is recommended.

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)

The paper has reported large changes in physiological variables at menopause which are mostly detrimental and that these are ameliorated with HRT (now called MHT in most countries)

I am not sure what is new in the conclusions as major studies have already shown that bone loss accelerates in relation to menopause, and with heavy bleeding women are more likely to have low ferritin/anemia and that this abates when bleeding ceases. Depressive symptoms only increase marginally in perimenopause and are still lower than in women aged less than 40 years where the prevalence of moderate to severe depressive symptoms is in the order of 30%.

My main concern is that the conclusions are limited by the source data
menopausal "stageing" is NHANES appears somewhat crude.

There is no mention of women who had an endometrial ablation or using a hormonal IUD which stops menstrual bleeding- these IUDs are common amongst women for contraception (14% of contraceptive users in 2015 in USA) and very common after 40 years for management of heavy bleeding/ as part of HRT.

There is something odd about the fraction of women reporting hysterectomy being 1%. In the US the frequency of hysterectomy has been reported as 19.2% at a mean age of 40.5 years. (doi:10.1016/j.maturitas.2013.08.007.) I am concerned why this % is so low

HRT data is also somewhat inconsistent with what has been published. Therefore,
I suggest great caution in the interpretation of the effects of menopausal hormone therapy (MHT) as 4% is extremely low usage even in the post WHI era.

HRT use according to NHANES data 1999-2020 showed a decline from ~37% to ~ 7% up to the age of 64 years which is very different from the 4% in this paper (JAMA Health Forum. 2024;5(9):e243128)

Notably, many women would not report being "on a progestin" if they had a hormonal IUD as they would not see this as the same thing. What about patches and gels?

Adjusting for sociodemographics and other variables does not adjust for unknown confounders- even if women are matched for a slew of variables, HRT users are intrinsically different to non users, especially when they are such a rarefied 4% group. This has led to over estimation of "benefits" of HRT in past studies that have been disproven in RCTs. So, suggesting HRT nullifies many of the effects of menopause is over interpreting these data. Thus I feel the claim that this study shows "Most of the physiological jumps are attenuated by hormone replacement therapy" is unsubstantiated by the data with only 4% using HRT and the fact that over and over again we have seen HRT users are different to non users due to a healthy user bias unexplained by simple sociodemographic characteristics.

I am interested in how the authors reconcile their data with the findings from Shen et al that consistent nonlinear patterns in molecular markers of aging, with substantial dysregulation occurs at two major points, namely at approximately 44 years and 60 years of chronological age. In both sexes. They reported that the transition point in women observed at approximately 55 years of age is not solely attributed to female menopause but, rather, represents a common phenomenon in the aging process of both sexes. (Nature Aging Aug 2024)

Comments regarding the introduction:

The following statement is not quite correct "This loss of estrogen and other feedback hormones of the hypothalamic-pituitary-ovarian axis causes follicle-stimulating hormone (FSH) and luteinizing hormone (LH) to rise sharply". The trajectories of both E and FSH around the FMP vary and I suggest this be acknowledged (European Journal of Preventive Cardiology, Volume 23, Issue 7, 1 May 2016, Pages 694–703, <https://doi.org/10.1177/2047487315607044>)

The way the introduction reports changes related to menopause also implies these changes occur after the final menstrual period (FMP) but the SWAN data has shown that dramatic changes are occurring from the early perimenopause.

I suggest the following needs modification to accurately represent what is known:

"the Study of Women's Health Across the Nation (SWAN), showing surges in arterial stiffness within a year of the FMP," ie endothelial dysfunction. SWAN showed "arterial stiffness" declines from the early perimenopause before estradiol levels have fallen, and the impairment in endothelial function does not accelerate after the FMP as this statement incorrectly suggests (see Moreau KL et al J Clin Endocrinol Metab, 2012 97(12):4692–4700).

Similarly if increased central fat was just due to estrogen deficiency it should accelerate after the FMP when estradiol is at its lowest, but it accelerates in the menopause transition when women are still making estradiol (Samargandy s et al Menopause. 2021 Mar 1;28(6):626–633)

I caution the implication that big data is better than a study like SWAN. the strength of SWAN is the impeccable documentation of the actual menopause transition in each participant. This is poorly documented in the big data sets and thus becomes more about age than menopausal status.

Finally, in the results:

I am not sure if the authors are referring to their study populations or globally (the ref is global) but the average age of menopause varies from ~47 in India to ~51.5 in Australia so that is not varying by 2 years as suggested.

(Remarks on code availability)

I am not a statistician- I am only able to comment on the data set which I believe is inappropriate for the analysis because of the limitations I have described. This cannot be fixed with clever statistics

Reviewer #3

(Remarks to the Author)

This paper examines the effects of menopause on the results of various common lab tests and the likelihood of developing a variety of age-related conditions. Experiments use a substantial (1.3M individuals) data set of lab-test results and diagnostic codes to inform the analysis. The authors deconvolve the trajectory of group-averaged biological function measurements against age with the estimated distribution of age-of-final-menstrual-period (FMP) to estimate the trajectory against time from FMP, which substantially sharpens the menopausal effect and highlights step changes.

In my view the deconvolution method is reasonable to highlight the effect qualitatively, but does not provide a clear quantitative estimates of menopause-induced change. The conclusions the authors draw largely align with this view, although occasionally they make quantitative statements, such as "Osteoporosis has a ten-fold rise in slope after menopause". I would suggest the deconvolution is not necessary to make the observation that menopause strongly increases the likelihood of osteoporosis, the precise effect of the deconvolution on the slope against age is a little hard to interpret so precise quantitative statements should be avoided, and perhaps the statement could be informed by data trajectories both before and after the deconvolution to give a more complete view.

In general the deconvolution idea may be a bit overemphasised in the paper. Many of the key conclusions about the effect of menopause on biological function and disease prevalence could be supported with the raw data, although the deconvolution does bring it into sharp relief. Another option would be to use the male cohort as a baseline against which to compare.

Overall however I do not have substantive methodological concerns and believe this paper makes some useful insights that confirm more anecdotal knowledge about the strong effect of menopause on biological function and disease prevalence.

(Remarks on code availability)

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

This study uses a large data set to confirm and expand on expected changes in the menopause transition. The authors have responded appropriately to the reviewers' concerns.

There are no real surprises and no mechanistic data. The cataloging of laboratory changes beyond what is available in SWAN and other studies is worthwhile, but the authors should be careful not to exaggerate the significance.

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)

I am happy with the changes the authors have made to address the concerns I raised. More broadly they appear to have addressed the concerns of the other reviewers carefully and comprehensively, although I cannot comment on the biological / medical discussions. I have no further comments on this manuscript.

(Remarks on code availability)

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

Our comments are in **black**, referee comments are in **blue**.

Reviewers' Comments:

Reviewer #1 (Remarks to the Author):

expertise: OBGYN, menopause

This manuscript presents a large-scale cross-sectional analysis of laboratory data to reconstruct physiological trajectories around the final menstrual period using a deconvolution approach. By analyzing more than 300 million laboratory measurements from two national cohorts, the authors report an abrupt, step-like change across multiple biological systems at menopause and suggest that these changes begin years before the final menstrual period and are attenuated among women using hormone replacement therapy.

The research question is important and the dataset is large, significant strengths of the manuscript.

We thank the reviewer for this endorsement.

However, **the manuscript is poorly organized and the language needs clarity**. The Results section is poorly structured and frequently includes methodological explanations, analytical rationale, and variable definitions that belong in the Methods section. The Results should be restricted to objective reporting of findings, without restating why analyses were performed, how variables were constructed, or providing interpretative statements.

Similarly, the Discussion is very brief and largely reiterates the results instead of interpreting them. There is little integration with prior literature and limited discussion of biological plausibility.

We thank the reviewer for this comment and have improved the organization and clarity of the manuscript. We revised the results, methods and discussion sections as suggested. We moved methodological explanations, analytical rationale and variable definitions to methods. We kept findings in the results, and moved interpretation to the discussion. The discussion now also includes integration with prior literature, discussion of biological plausibility and limitations.

The revised discussion now reads as follows:

“

We analyzed two large datasets with over 300 million laboratory tests, indicating that menopause is characterized by a "physiological cliff," defined as abrupt, step-like changes across nearly all physiological systems. We introduce a temporal deconvolution framework that uses cross-sectional data to generate a high-resolution timeline anchored to the FMP (final menstrual period). This can detect systemic shifts that are typically obscured by chronological age. The observed jumps at the FMP are concordant between the United States (NHANES) and Israel (Clalit) cohorts. The changes are most pronounced in systems with high estrogen sensitivity, such as bone metabolism, liver tests, lipid handling and osmolality. While prior

longitudinal studies, like SWAN, have established bone loss and specific hormonal surges, the scale of this study, covering 90 types of tests, demonstrates that this transition is sharp and systemic.

The magnitude of the jump in each test corresponds in part to the differential sensitivities of physiological systems to estrogen (Fig 2b). The large change of upstream hormones like FSH is due to their exquisite estrogen sensitivity via negative feedback in the hypothalamic-pituitary-ovarian axis. Bone is very sensitive to the protective effects of estrogen²⁴, followed by the liver⁸ which is a hub for lipid metabolism. Osmolality is regulated by antidiuretic (ADH) hormone (also called arginine vasopressin), whose sensitivity is governed by estrogen²⁵. Toxin levels such as lead also show a jump in menopause, probably due to the bone storage of lead²⁶. Several tests showed no detectable jumps at the FMP, including free thyroid hormones²⁷, PTH²⁸, folate²⁹ and fibrinogen³⁰. These systems are known to be weakly, indirectly, or debatably affected by estrogen and menopause. The jumps are preceded by accelerated changes in sex hormones more than a decade before menopause, earlier than the 3-7 year window traditionally ascribed²¹.

Notably, we find that abrupt physiological shifts are absent in males of the same age, suggesting that the cliff is a consequence of the menopausal transition rather than generalized aging. These lab test shifts are associated with a rise in risk for various clinical outcomes, which manifest in our analysis as a sudden increase in the slope of incidence at the FMP. Thus, prevalence is akin to an integral over time of the lab test trajectory. For instance, osteoporosis shows an increase in slope after menopause that strongly outpaces that of age-matched males, consistent with the large effects observed in bone-related laboratory tests. Cardiovascular outcomes, such as heart attack and stroke, also demonstrate an increased risk slope following the transition. Furthermore, we observed a transient rise in average depression scale ratings and sleep-related deficits around the time of the FMP, which then tend to decay back toward the slower aging patterns seen in males. Collectively, these findings suggest that the abrupt shifts in laboratory markers serve as a systemic inflection point for the accumulation of clinical health deficits.

Hormone replacement therapy, often referred to as menopausal hormone therapy (MHT), appeared to markedly attenuate the physiological shifts. In women reporting current HRT use, post-menopausal trajectories resembled smoother continuations of pre-menopausal states across various markers, including liver, renal, and bone tests. While HRT appeared to normalize most systems, it was associated with elevated CRP, an inflammation marker. This is consistent with studies that show that HRT estrogen stimulates hepatic CRP production in oral but not transdermal (patch) treatment due to a first pass hepatic effect^{31,32}. In general HRT is anti-inflammatory^{33,34}.

This study has several important limitations. It relies on retrospective cross-sectional data and self-reported menopause status. While deconvolution sharpens the timeline, it cannot replace the granularity of a prospective longitudinal study with detailed menopausal staging that tracks the same individuals through the transition. Adjusting for sociodemographics cannot entirely eliminate healthy user bias among HRT users, as those who choose or have access to HRT may differ intrinsically from non-users in ways not captured by medical records or demographic data. The effects of HRT need to be tested by prospective randomized trials. The accuracy of

our timelines also depends on the assumed distribution of FMP age, and while sensitivity analyses showed negligible effects to shifting this distribution, more accurate menopause tracking is warranted. Finally, while we controlled for many drugs, the lack of comprehensive data on hormonal IUDs or endometrial ablations in some cohorts could mask the true FMP for a subset of participants.

In conclusion, these findings further support a systemic and quantifiable view of the menopausal transition. We observed a physiological cliff at menopause across systems, which appears to be attenuated by hormone replacement therapy. This study extends the time frame for detectable changes and potential preventative care to almost a decade before menopause. More broadly, the present deconvolution approach can sharpen biological transitions from noisy, registry-scale data, a strategy transferable to other life-course events such as puberty, pregnancy and disease onset.

”

In addition, there is no section addressing the limitations of the study. Given the cross-sectional nature of the data, potential residual confounding, assumptions underlying the deconvolution approach, and uncertainty in the timing of the final menstrual period, a detailed limitations section is essential

We have now added a detailed limitation section addressing these points.

To the discussion we added as follows:

“

This study has several important limitations. It relies on retrospective cross-sectional data and self-reported menopause status. While deconvolution sharpens the timeline, it cannot replace the granularity of a prospective longitudinal study with detailed menopausal staging that tracks the same individuals through the transition. Adjusting for sociodemographics cannot entirely eliminate healthy user bias among HRT users, as those who choose or have access to HRT may differ intrinsically from non-users in ways not captured by medical records or demographic data. The effects of HRT need to be tested by prospective randomized trials. The accuracy of our timelines also depends on the assumed distribution of FMP age, and while sensitivity analyses showed negligible effects to shifting this distribution, more accurate menopause tracking is warranted. Finally, while we controlled for many drugs, the lack of comprehensive data on hormonal IUDs or endometrial ablations in some cohorts could mask the true FMP for a subset of participants.

”

The Methods section would also benefit from reorganization. A clearer structure would begin with a description of the datasets used, including references to their source protocols, followed by eligibility criteria, outcome and variable definitions, data processing steps, and finally the statistical analyses. This order would provide a more logical and transparent flow.

We have now reorganized the structure of the Methods section as suggested.

The methods section now reads:

“

Methods

Data and cohorts

We analyzed two datasets originating from different populations, USA (NHANES) and Israel (Clalit). NHANES is a cross-sectional health study that has been collected every 2 years from 1999-2022 that includes lab tests, health questions, exams and outcomes³⁵. Clalit is a large electronic medical record (all-comers) longitudinal dataset that includes about half the Israeli population¹⁵. We analyzed healthy, no meds lab test data that excluded confounding effects of medications and disease¹⁵, resulting in 1.3 million eligible women aged 30-70 from years 2004-2024.

Eligibility criteria

Clalit: We analyzed all comers aged 30-70 from years 2004-2024. We excluded pregnant women (N=173,000).

NHANES: We analyzed 12 biennial waves 1999-2022. We excluded pregnant women (N=1326). For HRT comparison (Figure 3) we only analyzed women from survey years 1999-2019 (2022 had no HRT data due to Covid), with known covariates needed for matching: menopause status, age, poverty income ratio, smoker status, insurance status, diabetes status, ethnicity and education, with final N=738 HRT users.

Variable definitions

Menopause status (NHANES only)

We included all non-pregnant women regardless of the reason for menopause (in sensitivity analysis we tried restricting to only natural menopause at typical ages, showing highly concordant results with Pearson's $r=0.97$, SI Fig S7). Menopause status (females only) was determined from the clinical definition of 12 months without period² together with supplemental information to confirm status. We used two questions, the first is: have you “had regular/at least one period in the past 12 months?”. The second follow-up question is “what is the reason (for no periods)?”, where we require the individual to specify “menopause” as the reason. Individuals that self-reported having periods were defined as non-menopausal (RHQ030/RHQ031: regular periods/at least one period in the last 12 months). Individuals were further classified as follows. Those self-reporting not having periods were classified as non-menopausal if self-reported pregnant (this year or previous) or breastfeeding, whereas individuals who self-reported menopause were classified as menopausal (RHQ040/RHD042/RHD043: reason not having regular periods); individuals that reported any other reason for not having periods were classified as unknown (NA) menopause status. RHD042 did not differentiate menopause from hysterectomy as the reason for not having regular periods but these cases were corrected afterwards as follows. We reclassified all women who had a hysterectomy that reported having either a single ovary or no ovaries removed as unknown (NA) menopause age and status, since the hysterectomy would prevent them from having periods (RHQ300/RHQ305/RHQ310/RHQ320); women who reported their menopause age before their

hysterectomy age (RHQ291) were still included. We included participants taking oral contraceptives. Some of these individuals may not have periods in which case they were classified as unknown menopausal status. Finally, some individuals with indeterminate menopause status were classified according to age as follows: non-responders to regular periods (RHQ030/RHQ031) under the age of 30 were classified as non-menopausal, and responders over the age of 60 who reported not having regular periods were classified as menopausal (justified by the observed statistics that showed 98.3% of women are non-menopausal at age 30 and only 0.9% of women were non-menopausal at age 60). Self-reported age of menopause is top-coded at 60 in NHANES, that was extrapolated using survival modelling³⁶ as described in Methods subsection YoGlen Method 1. Clalit did not have menstrual data so menstrual status was not used for that dataset. Males and the 26.8% of women with unknown menopause status were imputed with a menopause age using the YoGlen algorithm described later in the methods section. We performed sensitivity analysis to consider the effects of unknown menopause status, hysterectomy and early/late menopause in SI Fig S7, that shows nearly perfect concordance of results, $r=0.97$, between a restricted age range of 45-53 years old with known menopause status and our reported FMP jump sizes.

Hormone replacement therapy (HRT)

NHANES: We considered females as currently taking HRT if they had known post-menopause status and either (i) self-reported taking HRT or (ii) reported taking any drugs containing estrogens (full list in SI Table S5). For (i) we gated NHANES using RHQ540 (ever used female hormones excluding birth control): any participant reporting “no” was classified as no HRT. Affirmative to any of the following was classified as yes HRT: RHQ558 (taking estrogen-only pills now; 2.9% of females), RHQ566 (taking progestin-only pills now; 0.2% of females), or RHQ574 (taking combo pill now; 1.1% of females). After 2012 NHANES stopped asking specific questions regarding HRT usage. For (ii) we used the RQQ_RX prescription drug files from each year except 2022 (not reported). We included the following types of estrogen-containing drug classes as HRT: all “estradiol”, estropipate, mestranol, all “estrogen”, as well as 9 individuals using progesterone-only HRT.

Clalit: In Clalit, menopausal status is not recorded. We therefore defined HRT users as women aged 45–55 who were prescribed at least twice the HRT drugs Angeliq (drospirenone 2 mg / estradiol hemihydrate 1 mg) or Eviana (estradiol hemihydrate 0.5 mg / norethisterone acetate 0.1 mg). We matched these individual users with controls from the same clinic, age and birth year. For each laboratory test and each group, we defined the menopause change Δ as the mean of the post-menopausal window (ages 54–60) minus the corresponding mean in the pre-menopausal window (ages 42–48). Δ_{HRT} is the mean menopause change for HRT patients, $\Delta_{control}$ is the mean menopause change for the control group. HRT attenuation is defined as the relative reduction in the menopause change of HRT users compared with their matched controls: $1 - (\Delta_{HRT} / \Delta_{control})$.

Variables and Outcomes

We analyzed 90 lab blood tests and 31 outcomes for their time dependence with respect to inferred time of FMP. We group by systems as described in SI Table S1,S4. Blood tests describe

bone, liver, kidney, immune and other systems. Outcomes include signs, symptoms, and diagnoses.

Preprocessing

Clalit and NHANES: All lab values were log scaled and standardized to sex-specific mean and standard deviation using the control reference age range of 25-45 (datasets were standardized separately). An outlier cutoff was applied at ± 5 STDs (<0.3% of datapoints). Log lab value distributions resembled normal distributions.

Clalit: Lab values were filtered to exclude measurements from non-healthy individuals and from individuals taking medications that affect each specific lab test. To do so, we used the dataset described by Cohen et al¹⁵, which excluded 131 chronic conditions and 5,223 drug-test pairs as well as pregnancies. BMI was linked cross-sectionally for the healthy population using the most recent measurement within the study period, and reported as median [IQR]. Bootstrapping on Clalit was performed by grouping individuals by year of test date and sampling the 20 years with replacement. Test values of zero were removed (<0.3% of datapoints, SI Table S2).

The YoGlen Algorithm

We estimate the unknown age of FMP using a latent variable model³⁷ with appropriate approximations. Let p denote an arbitrary probability density function, a_m is the unknown age of FMP, m is the binary menopause status variable (available in NHANES but not Clalit), a is the known age, and $\vec{y}$ is a vector of known lab test values. The two results we derive are (1) how to estimate the distribution of the unknown age of FMP of each individual, $p(a_m | \vec{y}, a)$ using the observed data, and (2) how to estimate mean cross-sectional lab trajectories as a function of the unknown time to FMP, $E[y_j | a - a_m]$. Jump size was estimated using resampled regression model predictions. For Method 1 we used 10 repeat, 10 fold cross-validation. For Method 2 we used bootstrapping of $E[y_j | a - a_m]$.

YoGlen Method 1 for NHANES data

To estimate the unknown FMP time, $p(a_m | \vec{y}, a)$ we use Bayes' theorem

$p(a_m | \vec{y}, a, m) \propto p(a_m | a, m) p(\vec{y} | a_m, a, m)$. We observed based on comparison to SWAN data (SI Fig S5) that a good approximation was to use only the lowest order term

$p(a_m | \vec{y}, a, m) \approx p(a_m | a, m)$. We used draws from $p(a_m | a, m)$ to impute the unknown a_m where $p(a_m)$ is the reference menopause age distribution from NHANES (Fig 1), estimated by

modelling the log-hazard as a 10-knot natural spline³⁶. Women with double oophorectomies (12%) were drawn separately using their own distribution. For females where menopause status m is known (73.2%), conditioning on a and m restricts the permitted a_m to be before current age if non-menopausal and after current age if menopausal. We used a discrete grid to consider all possible a_m from 20-70 at 1 year resolution. Once the $p(a_m | a, m)$ grid is estimated for each

individual, we use it to draw 15 possible a_m for each individual, creating 15 completed datasets that are analyzed in parallel and then the results are pooled using Rubin's rules³⁸ to account for the uncertainty in a_m .

In NHANES, most (89%) of the self-reported post-menopausal females also included their age of menopause. Prior research indicates that self-reported age of menopause is accurate to about 1 year³⁹, and hence age of FMP for these participants was estimated assuming $p(a_m)$ was a normally distributed true age of menopause with mean equal to their reported value and a 1 year standard deviation. Validation by comparison to longitudinal data from SWAN demonstrates that YoGlen fits well to the population mean for the key variables available in SWAN, FSH and E2, when only age and menstrual status are included (SI Fig S5). Inclusion of FSH and E2 into the model tended to fit worse (not shown), ostensibly because of the strong heteroscedasticity due to fluctuations in sex hormones that occurs during perimenopause^{21,40} and suddenness of the transition. Therefore for NHANES the overall population trends were estimated using only menstrual status and age.

YoGlen Method 2 for Clalit data

For Clalit we did not have access to menstrual status data, precluding using YoGlen Method 1, so we instead used a regression model to deconvolve each lab test to convert the age-dependent cross-sectional means into time-since-FMP dependent dynamics, effectively using $p(a_m | y_j, a, m) \propto p(a_m | a) p(y_j | a_m, a)$. We modelled each lab test as piecewise linear,

$$f(t|a) \equiv \sum_i \alpha_i \cdot (t - t_i) I(t \geq t_i \& t < t_{i+1}) + \beta_i, \text{ where } t \equiv a - a_m \text{ is the time to menopause, } I \text{ is}$$

the indicator function (1 when true, 0 otherwise), and α_i and β_i are the regression model parameters. All models were separately fit for pre ($t < 0$) and post ($t > 0$)-FMP, permitting a jump. Averaging this equation over age a yields the cross-section mean as function of time t to FMP

$$E[y|t] = \sum_i \alpha_i \cdot (E[t] - t_i) I(t \geq t_i \& t < t_{i+1}) + \beta_i.$$

where $E[t] = a - E[a_m]$ and $E[a_m]$ were computed from a published menopause age distribution from 306,068 healthy females¹⁷.

Propensity score matching (PSM) was used to match HRT recipients after menopause to controls both before and after menopause (1:1 matching ratio). Propensity score matches each individual to similar controls using logistic regression on a set of covariates to determine similarity²⁰. For controls we matched to: age, poverty income ratio > 1 (yes/no), smoker status (smokes now/does not), insurance status (has/does not), diabetes status (positive/negative), age of menopause, survey year ≥ 2005 (binary), black ethnicity (yes/no) and education $\geq$ "some college" (yes/no). Black ethnicity was included as black females had much higher BMD than non-black (U-test $p=10^{-16}$, mean difference: 0.75 ± 0.02 z-score total BMD). Score balance was determined both visually using absolute mean difference, and KS threshold of 0.1.⁴¹ All covariates were balanced in the after menopause control group, but age differed in the before

menopause control. Pre-menopause control group differed by their median (IQR) of 51 years (48-53 years) versus 61 years (54-67 years) for the HRT group; this was expected and tolerated since the HRT are post-menopausal and therefore typically older.

Regression fit lines The regression lines in Figs 1, 2, 4 and 5 were computed using model averages weighted by repeated cross validation. To do so we performed model selection (models listed below) using 10-fold, 10-repeat cross-validation. We used the out-of-sample mean-squared error to choose continuous models, and AUC⁴² (area under the ROC curve) to choose binomial models. Each fold included fitting the model to each of the FMP age estimates from YoGlen then pooling using Rubin's rules³⁸; pooled individual predictions were used to compute the error. In Figs 1 and 2 we present the model average where we picked one of 4 Gaussian regression models at each repeat (10-knot spline, 10-knot spline with jump, 10-knot location-scale, or piecewise 10-knot location-scale) then aggregated grid predictions from -25 to 25 years-to-FMP over the 100 total repeats. The location-scale models included joint prediction of mean and variance to mitigate issues with heteroskedasticity, likely originating from late menopause transition fluctuations²¹ (e.g., Fig 4 g-i). Similarly, in Fig 5 we present the model average where we picked one of 2 weighted binomial regression models at each repeat because the outcome variables are binary and location-scale models are therefore not relevant (10-knot spline, 10-knot spline with jump). Rare outcomes risk underfitting and therefore we set binomial case weights to 1 for the majority outcome and majority / minority ratio for minority outcome⁴³. In Fig 4 we present data for females with known non-menopausal status, and therefore did not require model selection. We present the 10-knot Gaussian location-scale model which allows estimates of STD. Best fit lines used the `mgcv` package in R⁴⁴. In analysis of HRT (Fig 3) we used linear regression on data pooled using Rubin's rules due to the relatively small number of data points.

Statistics

Statistics were performed in R v4.5.1 (NHANES) and Python v3.9.23 (Clalit). All error bars/bands are standard errors. Correlations are Pearson's. All tests are two-sided. Significance was set at $p < 0.05$. False discovery rate (FDR) correction used the Benjamini-Hochberg algorithm⁴⁵.

All methods are available on GitHub: <https://github.com/AlonLabWIS/YoGlen>.

”

The conclusion overstates the implications of the findings. Phrases such as “these findings redefine menopause as a quantifiable, systemic transition” imply a level of conceptual redefinition that is not fully supported. A more appropriate wording would be that the findings may contribute to or “further support” a systemic and quantifiable view of the menopausal transition, rather than claiming a definitive redefinition.

We have reduced our claims as suggested, specifically including in the abstract and discussion section.

The abstract now reads:

“

Menopause fundamentally reshapes female physiology, yet current understanding is limited by small longitudinal cohorts that characterize it as a gradual transition. Large-scale biomedical datasets remain underutilized because the age of the final menstrual period (FMP) is rarely recorded. Here, we present a computational framework that leverages cross-sectional data to reconstruct systemic physiology as a function of time relative to FMP. We adapted a deconvolution framework from astronomy to recover systemic biological trajectories by deconvolving the population distribution of FMP age from chronological data. Applying this to two national cohorts with 300 million laboratory tests from 1.3 million women, we transformed cross-sectional measurements into high-resolution timelines anchored to the FMP. Our analysis reveals a distinct, step-like physiological cliff at the FMP across endocrine, skeletal, hepatic, renal, inflammatory, and lipid systems. These discontinuities are absent in males and highly concordant across independent populations. We demonstrate that systemic dysregulation begins over a decade prior to FMP, significantly expanding the window for preventive intervention. Furthermore, we show that hormone replacement therapy (HRT) markedly attenuates these abrupt physiological shifts. These findings further support a systemic and quantifiable view of the menopausal transition and provide a generalizable strategy for recovering hidden biological trajectories from human datasets, applicable to other life stages such as puberty or disease progression.

”

The conclusion now reads:

“

In conclusion, these findings further support a systemic and quantifiable view of the menopausal transition. We observed a physiological cliff at menopause across systems, which appears to be attenuated by hormone replacement therapy. This study extends the time frame for detectable changes and potential preventative care to almost a decade before menopause. More broadly, the present deconvolution approach can sharpen biological transitions from noisy, registry-scale data, a strategy transferable to other life-course events such as puberty, pregnancy and disease onset.

”

The authors should clarify how they handled women using hormonal medications, such as oral contraceptives, who may not menstruate but are not menopausal.

We included participants taking oral contraceptives. In NHANES, we used two questions, the first is: have you “had regular/at least one period in the past 12 months?”. The second follow-up question is “what is the reason (for no periods)?” We did not mark as menopausal individuals who answered the second question with any reason other than menopause (except for the

hysterectomy edge case described below). Individuals not menstruating due to contraceptives are treated as unknown menopausal status.

We now detail this in methods that reads:

“

Menopause status (NHANES only)

We included all non-pregnant women regardless of the reason for menopause (in sensitivity analysis we tried restricting to only natural menopause at typical ages, showing highly concordant results with Pearson's $r=0.97$, SI Fig S7). Menopause status (females only) was determined from the clinical definition of 12 months without period² together with supplemental information to confirm status. We used two questions, the first is: have you “had regular/at least one period in the past 12 months?”. The second follow-up question is “what is the reason (for no periods)?”, where we require the individual to specify “menopause” as the reason. Individuals that self-reported having periods were defined as non-menopausal (RHQ030/RHQ031: regular periods/at least one period in the last 12 months). Individuals were further classified as follows. Those self-reporting not having periods were classified as non-menopausal if self-reported pregnant (this year or previous) or breastfeeding, whereas individuals who self-reported menopause were classified as menopausal (RHQ040/RHD042/RHD043: reason not having regular periods); individuals that reported any other reason for not having periods were classified as unknown (NA) menopause status. RHD042 did not differentiate menopause from hysterectomy as the reason for not having regular periods but these cases were corrected afterwards as follows. We reclassified all women who had a hysterectomy that reported having either a single ovary or no ovaries removed as unknown (NA) menopause age and status, since the hysterectomy would prevent them from having periods (RHQ300/RHQ305/RHQ310/RHQ320); women who reported their menopause age before their hysterectomy age (RHQ291) were still included. We included participants taking oral contraceptives. Some of these individuals may not have periods in which case they were classified as unknown menopausal status. Finally, some individuals with indeterminate menopause status were classified according to age as follows: non-responders to regular periods (RHQ030/RHQ031) under the age of 30 were classified as non-menopausal, and responders over the age of 60 who reported not having regular periods were classified as menopausal (justified by the observed statistics that showed 98.3% of women are non-menopausal at age 30 and only 0.9% of women were non-menopausal at age 60). Self-reported age of menopause is top-coded at 60 in NHANES, that was extrapolated using survival modelling³⁶ as described in Methods subsection YoGlen Method 1. Clalit did not have menstrual data so menstrual status was not used for that dataset. Males and the 26.8% of women with unknown menopause status were imputed with a menopause age using the YoGlen algorithm described later in the methods section. We performed sensitivity analysis to consider the effects of unknown menopause status, hysterectomy and early/late menopause in SI Fig S7, that shows nearly perfect concordance of results, $r=0.97$, between a restricted age range of 45-53 years old with known menopause status and our reported FMP jump sizes.

”

In addition, if women under 30 years who did not respond to questions on regular menstruation were classified as non-menopausal, the choice of this age threshold should be justified, either by rationale or by citation of prior literature.

We added a rationale for the age 30 threshold in Methods section as follows:

“

Finally, some individuals with indeterminate menopause status were classified according to age as follows: non-responders to regular periods (RHQ030/RHQ031) under the age of 30 were classified as non-menopausal, and responders over the age of 60 who reported not having regular periods were classified as menopausal (justified by the observed statistics that showed 98.3% of women are non-menopausal at age 30 and only 0.9% of women were non-menopausal at age 60)...We performed sensitivity analysis to consider the effects of unknown menopause status, hysterectomy and early/late menopause in SI Fig S7, that shows nearly perfect concordance of results, $r=0.97$, between a restricted age range of 45-53 years old with known menopause status and our reported FMP jump sizes.

“

Since NHANES provides a direct definition of menopause whereas Clalit relies on inferred menopausal timing, the credibility of the results would be strengthened by a sensitivity analysis restricted to NHANES participants with directly defined menopausal status.

This comment helped us add new sensitivity analyses. We restricted the analysis to individuals from NHANES with known menopause status in the age-range 45-53 years who did not have a hysterectomy. We also test two alternative algorithms to YoGlen. We use both the population mean difference and propensity-score matched pair difference. In all cases we find very high concordance ($r \geq 0.97$) between the estimated changes reported in Fig 1 and all sensitivity analyses. We now detail these sensitivity tests in the new SI section “Sensitivity analysis” (Fig S7).

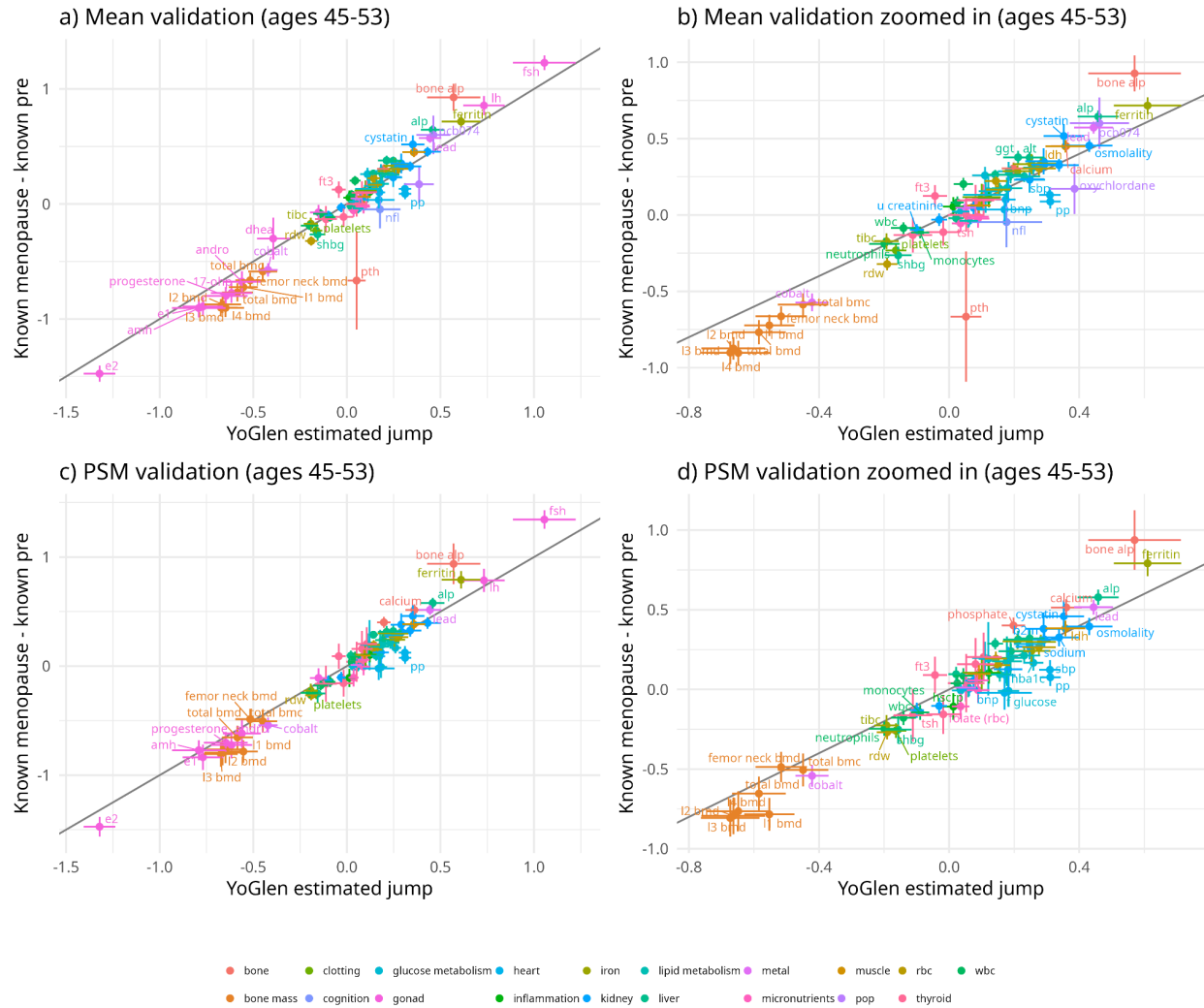


Fig S7. Sensitivity analysis of jump sizes shows strong concordance between YoGlen and direct estimates (NHANES). We compared the difference between pre-menopausal and post-menopausal women aged 45-53 who did not have a hysterectomy with known menopause status (y-axis) to the YoGlen estimates from Figure 2 (x-axis), both with (left) and without (right) the sex hormones. **a)** We see strong concordance between the mean difference and YoGlen estimate, with Pearson's rho = 0.97 (0.95-0.98 95% CI, $p=10^{-16}$; $n=3155$ individuals). **b)** Same as figure a) but zoomed in (i.e., excluding sex hormones). **c)** We see strong concordance with the mean PSM differences and the YoGlen estimate. Pearson's rho = 0.98 (0.97-0.99 95% CI, $p=10^{-16}$, $n=890$ pairs). **d)** Same as figure c) but zoomed in. Known menopause status only. PSM is paired difference after matching based on: poverty, smoking, insurance, diabetes, pre vs post 2005, black ethnicity and "some college" (or more) education level (all balanced, with absolute mean difference and KS threshold of 0.1⁴¹). Minimum 50 measurements per test.

The current title does not specify the study design. From the title alone, it is unclear what type of study was conducted.

We have incorporated the study design into the title, the new title is “Systemic physiological cliff at menopause revealed by temporal deconvolution of 300 million lab tests: a multi-cohort retrospective study”.

In the Introduction and Discussion, several statistics and factual statements are cited from review articles rather than from the original studies that first reported these findings. For example, the statement “more than 1.1 billion women are living beyond their final menstrual period” is referenced through secondary sources.

We have now removed the 1.1 billion sentence. We added primary references to the introduction that now includes the following original studies:

1. Ruth, K. S. et al. Genetic insights into biological mechanisms governing human ovarian ageing. *Nature* 596, 393–397 (2021).
4. Tepper, P. G. et al. Trajectory clustering of estradiol and follicle-stimulating hormone during the menopausal transition among women in the Study of Women’s Health across the Nation (SWAN). *J. Clin. Endocrinol. Metab.* 97, 2872–2880 (2012).
5. Randolph, J. F., Jr et al. Change in follicle-stimulating hormone and estradiol across the menopausal transition: effect of age at the final menstrual period. *J. Clin. Endocrinol. Metab.* 96, 746–754 (2011).
9. Samargandy, S. et al. Arterial stiffness accelerates within 1 year of the final menstrual period: The SWAN heart study: The SWAN heart study. *Arterioscler. Thromb. Vasc. Biol.* 40, 1001–1008 (2020).
10. Samargandy, S. et al. Abdominal visceral adipose tissue over the menopause transition and carotid atherosclerosis: the SWAN heart study: the SWAN heart study. *Menopause* 28, 626–633 (2021).
12. Bromberger, J. T. et al. Major depression during and after the menopausal transition: Study of Women’s Health Across the Nation (SWAN). *Psychol. Med.* 41, 1879–1888 (2011).
13. Mrowiec, K. et al. Association of serum metabolome profile with the risk of breast cancer in participants of the HUNT2 study. *Front. Oncol.* 13, 1116806 (2023).
14. Richardson, W. H. Bayesian-based iterative method of image restoration. *J. Opt. Soc. Am.* 62, 55 (1972).

Finally, abbreviations are not used consistently. Some terms are abbreviated before being written in full, and in other cases the full term and abbreviation are reintroduced after already being defined. Consistent use of abbreviations throughout the manuscript is recommended.

We have reviewed all abbreviations for consistency and revised accordingly.

Reviewer #2 (Remarks to the Author):

expertise: menopause, women's health

The paper has reported large changes in physiological variables at menopause which are mostly detrimental and that these are ameliorated with HRT (now called MHT in most countries)

Regarding the term MHT, we have updated the results and discussion section to refer the reader to this updated nomenclature as follows.

In the results we write: "We tested the effect of hormone replacement therapy (HRT, often called menopausal hormone therapy MHT)"

In the discussion we write: "Hormone replacement therapy, often referred to as menopausal hormone therapy (MHT),"

I am not sure what is new in the conclusions as major studies have already shown that bone loss accelerates in relation to menopause, and with heavy bleeding women are more likely to have low ferritin/anemia and that this abates when bleeding ceases. Depressive symptoms only increase marginally in perimenopause and are still lower than in women aged less than 40 years where the prevalence of moderate to severe depressive symptoms is in the order of 30%.

We now better emphasize that major studies such as SWAN have already established the acceleration in bone loss and some other results.

In the introduction we write: "The menopause transition was illuminated by prospective cohorts such as the Study of Women's Health Across the Nation (SWAN)^{5,9-12}, showing that sex hormone changes begin about 7 years before menopause⁵, and that bone loss accelerates in relation to menopause¹¹. Such longitudinal studies are crucial due to their exquisite menopause documentation."

In the results we credit previous work: "The latter tests improve (move in the direction of less anemia) probably due to the cessation of menstrual bleeding¹⁹."

We also clarify the novelty of this study: a method that opens up the use of large retrospective datasets to study biological transitions using deconvolution. This allows use of a large number of participants (1.3M) and a comprehensive range of tests (90 types of tests) to find transitions across major organ systems. Such scale and scope is hard to do in prospective studies. The finding of abrupt change in many systems at the FMP sharpens previous longitudinal studies such as SWAN which had a 2-year resolution and considered fewer tests and individuals. Our method allows following data across several decades to uncover the long tail of menopause changes, and the early detectable rise on a population level. The matching between two populations and study designs (US and Israel, cross-sectional study and all-comers medical records) strengthens generality, although the gold standard remains longitudinal studies with accurate menopause staging.

My main concern is that the conclusions are limited by the source data menopausal "stageing" is NHANES appears somewhat crude.

We now emphasize a comparison of our deconvolution approach with the SWAN longitudinal data average aligned with the FMP. We find good agreement ($r \geq 0.98$) for the key tests that overlap with NHANES, E2 and FSH. We show this in SI Fig S5. We also mention in the limitation section “This study has several important limitations. It relies on retrospective cross-sectional data and self-reported menopause status. While deconvolution sharpens the timeline, it cannot replace the granularity of a prospective longitudinal study with detailed menopausal staging that tracks the same individuals through the transition.”

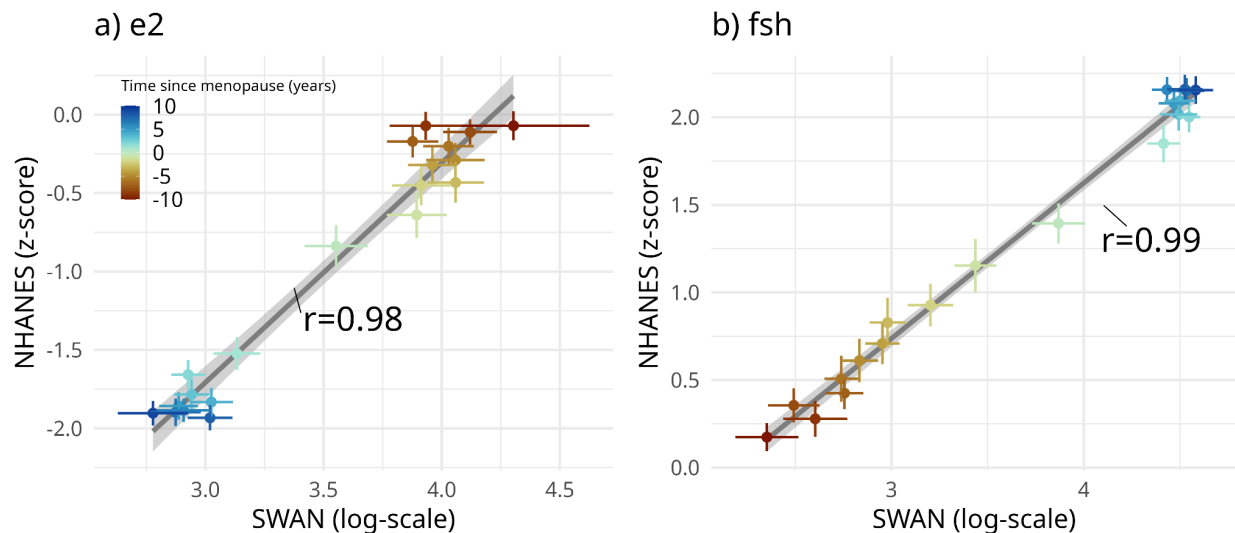


Fig S5. Validation of YoGlen method (NHANES) by comparison to the longitudinal Study of Women's Health Across the Nation (SWAN) for the key sex hormones. a) e2 (Pearson's $r=0.98$, $p=10^{-14}$). b) fsh ($r=0.99$, $p<10^{-15}$). 1-year bins from -10 to 10 years before/after menopause. SWAN data are log-scaled; NHANES data are log-scaled then z-scored by healthy reference (Table S1).

There is no mention of women who had an endometrial ablation or using a hormonal IUD which stops menstrual bleeding- these IUDs are common amongst women for contraception (14% of contraceptive users in 2015 in USA) and very common after 40 years for management of heavy bleeding/ as part of HRT.

We tested sensitivity to IUD and other conditions that cause lack of menstruation which is not natural menopause by repeating our analysis using only participants who specifically answered two questions: the first is lack of menstruation in the past years, and the second follow-up questions is “why?” with the answer “menopause” or for some waves it was “menopause or hysterectomy”. From these latter cases, we removed all participants with hysterectomy. We find that the results in this cohort are nearly identical to the analysis in the paper ($r \geq 0.97$). We now detail this in the new SI Sensitivity analysis section, specifically Fig S7.

We also write in the limitations section “Finally, while we controlled for many drugs, the lack of comprehensive data on hormonal IUDs or endometrial ablations in some cohorts could mask the true FMP for a subset of participants.”

The SI Sensitivity analysis section now includes the following:

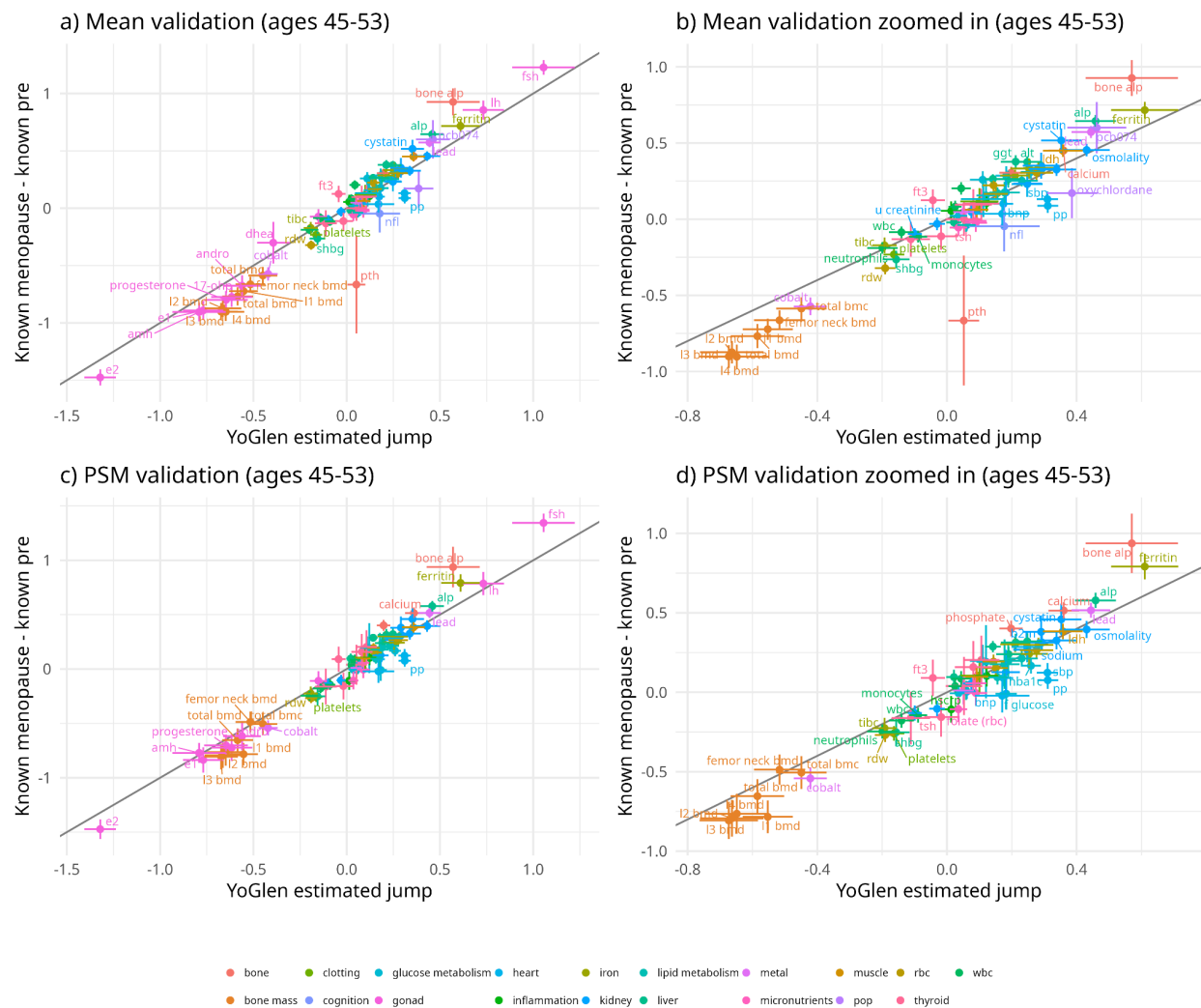


Fig S7. Sensitivity analysis of jump sizes shows strong concordance between YoGlen and direct estimates (NHANES). We compared the difference between pre-menopausal and post-menopausal women aged 45-53 who did not have a hysterectomy with known menopause status (y-axis) to the YoGlen estimates from Figure 2 (x-axis), both with (left) and without (right) the sex hormones. **a)** We see strong concordance between the mean difference and YoGlen estimate, with Pearson's rho = 0.97 (0.95-0.98 95% CI, $p=10^{-16}$; $n=3155$ individuals). **b)** Same as figure a) but zoomed in (i.e., excluding sex hormones). **c)** We see strong concordance with the mean PSM differences and the YoGlen estimate. Pearson's rho = 0.98 (0.97-0.99 95% CI, $p=10^{-16}$, $n=890$ pairs). **d)** Same as figure c) but zoomed in. Known menopause status only. PSM is paired difference after matching based on: poverty, smoking, insurance, diabetes, pre vs post 2005, black ethnicity and "some college" (or more) education level (all balanced, with absolute mean difference and KS threshold of 0.1⁴¹). Minimum 50 measurements per test.

There is something odd about the fraction of women reporting hysterectomy being 1%. In the US the frequency of hysterectomy has been reported as 19.2% at a mean age of 40.5 years. (doi:10.1016/j.maturitas.2013.08.007.) I am concerned why this % is so low

We thank the reviewer for this comment that helped us improve our inclusion criteria. We have updated our criteria and no longer mention these 1%. To clarify, the 1% estimate in the manuscript was not women reporting hysterectomies in general, but rather women reporting hysterectomies that were not menopausal. We observed that 25% of NHANES women reported hysterectomies, but only 9% had hysterectomies without double oophorectomies, and of these the 1% mentioned in the paper were those at pre-menopausal age.

To address this, we now set the menopause status of women who had hysterectomy that did not cause menopause (i.e. they did not have a double oophorectomy) to “menopause status unknown”. We have since verified that our updated results are nearly identical to our initial analysis (Pearson’s $r=1.00$, 95% CI: 0.99-1 for jump size).

HRT data is also somewhat inconsistent with what has been published. Therefore, I suggest great caution in the interpretation of the effects of menopausal hormone therapy (MHT) as **4% is extremely low usage** even in the post WHI era.

HRT use according to NHANES data 1999-2020 showed a decline from ~37% to ~7% up to the age of 64 years which is very different from the 4% in this paper (JAMA Health Forum. 2024;5(9):e243128). Notably, many women would not report being “on a progestin” if they had a hormonal IUD as they would not see this as the same thing. What about patches and gels?

This comment helped us to enlarge our cohort of HRT users. Previously we used only self-reported “current use of HRT”. In the revised manuscript we now include post-menopausal women taking estrogen with or without progestin (SI Table S5). This raises the fraction using HRT to 9%. We also added new survey years (2012-2019).

The results remain consistent with the previous version, despite having three times more HRT participants, as shown in the new SI Fig S8 ($r=0.90$).

We also performed an analysis of HRT use in Clalit, that was absent from the previous manuscript. We find concordance between the two cohorts. In Clalit, we used a longitudinal design, where we compared individuals prescribed Eviana(Estradiol as hemihydrate 0.5 mg + Norethisterone acetate 0.1 mg) or Angeliq (Estradiol (as hemihydrate) 1mg + Drospirenone 2mg) tablets at least twice, to individuals of the same birth year and age from the same clinic (clinics average $xx \pm xx$ females and represent localities often with similar socioeconomic properties as documented in ref xxx]. The concordance of NHANES and Clalit is moderate-high ($r=0.63$ $p=0.005$).

Thirdly, we compared NHANES propensity score matching (which has $n \sim 1300$) to a distinct control approach of regression of a tenfold larger population aged 40-70 ($n \sim 13,000$) . The concordance of the two methods is excellent ($r=0.90$, $p=10^{-16}$).

Regarding patches and gels, we analyzed Clalit data for patches (evolrel xxx). We do not have enough data for statistically meaningful analysis.

We have updated Figure 2 to include the new results:

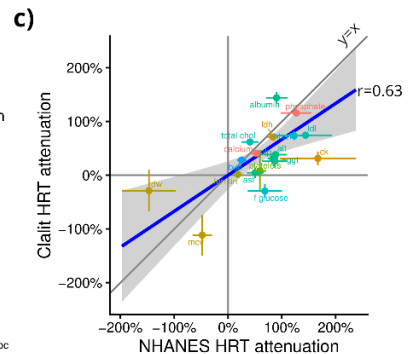
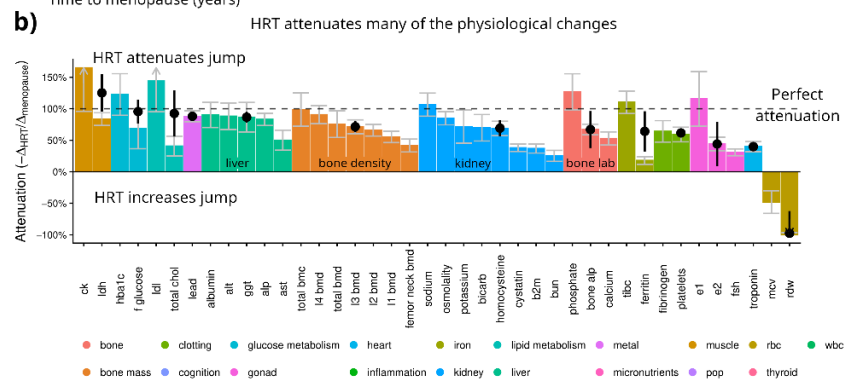
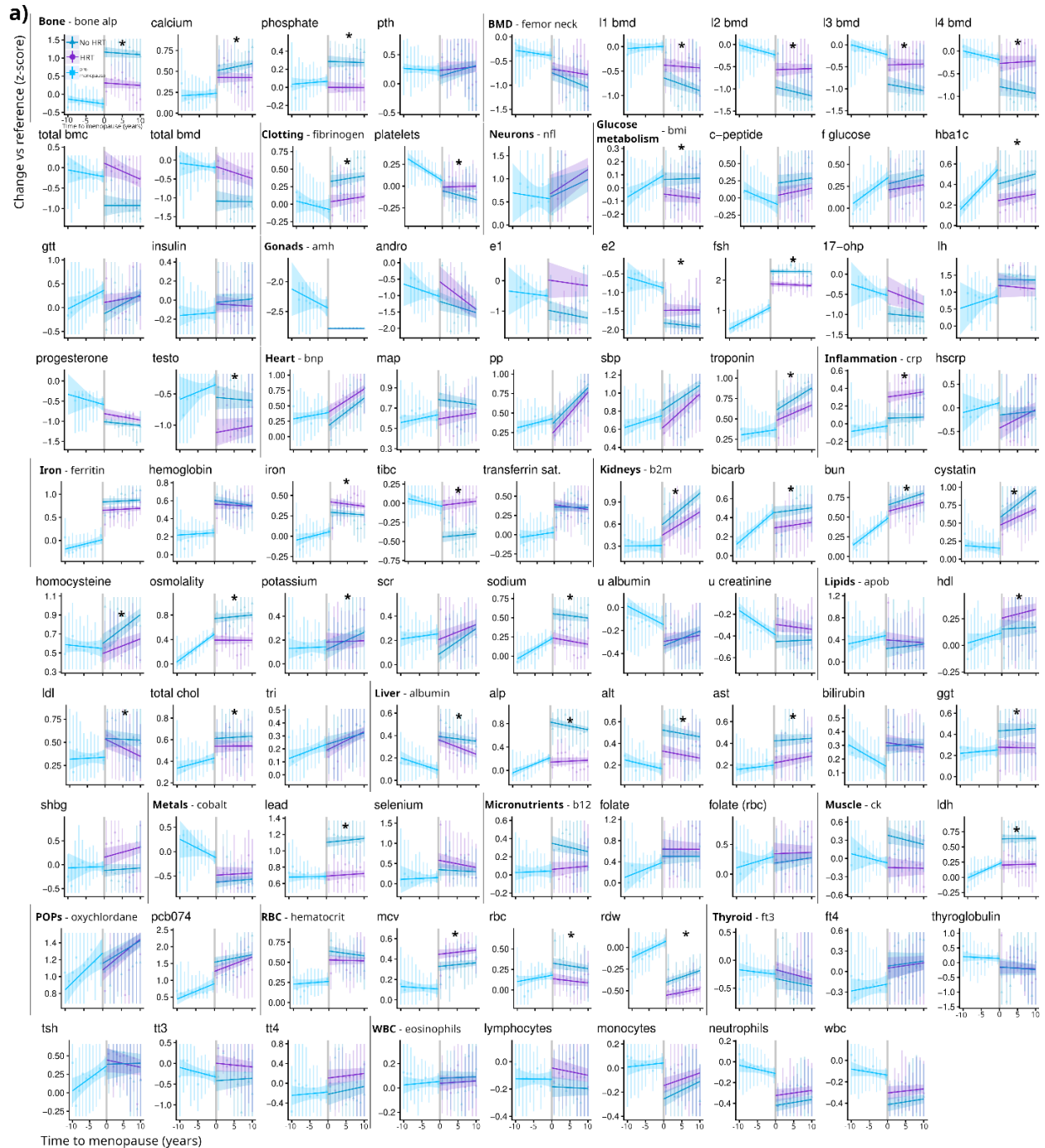


Fig 3. Hormone replacement therapy (HRT) appears to reduce jumps in many systems. **a)** NHANES lab tests as a function of time to menopause comparing participants with current HRT use (purple) to propensity score matched controls (blue; median n = 378 IQR: 114-710 with test and taking HRT, see SI Table S3); matched premenopause participants also in blue. Star marks significant HRT effects (FDR corrected paired Mann–Whitney U test $p < 0.05$). **b)** Jump attenuation defined as the ratio of jumps in HRT versus control, $(\text{HRT-pre})/(\text{control-pre})$, shown are significantly changed tests (FDR corrected unpaired z-test $p < 0.05$). 100% attenuation indicates complete abolition of the jump at the final menstrual period (FMP), negative values means a larger jump in the HRT cohort. Tests are grouped by system, black points are the median $\pm$ mean absolute deviation of the system. All lab tests have at least 10 measurements for each group (HRT now, no HRT, and premenopause). **c)** NHANES and Clalit show moderate-strong agreement on HRT attenuation for the overlapping variables from b), ($r=0.63$, $p = 0.005$; including only Clalit variables with >200 individuals, Methods).

We have added to the supplemental:

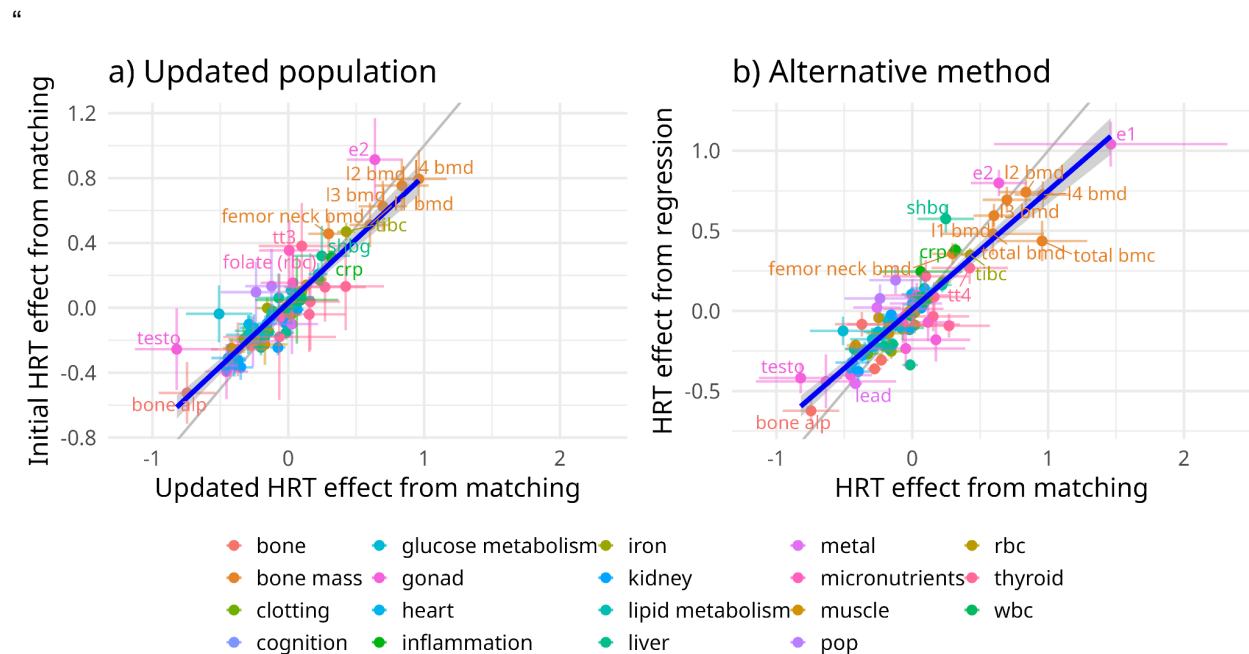


Fig S8. Sensitivity analysis of HRT effect sizes indicates that the results are robust to changing the population (a) and changing the estimating algorithm (b). **a)** Our initial effect sizes estimated using self-reported HRT status from years 1999-2012 only (y-axis) correlated strongly with our updated estimates including specific drugs used for years 1999-2019 (x-axis), $r=0.90$ (95% CI: 0.85-0.93, $p=10^{-16}$); propensity-score matched. **b)** Estimates using ordinary linear regression (y-axis) correlate strongly with our updated estimates $r=0.90$ (95% CI: 0.86-0.94, $p=10^{-16}$). Propensity-score matched using: survey period (pre vs post 2005), menopause status, age, poverty income ratio, smoker status, insurance status, diabetes status, black ethnicity (updated only) and education. Ordinary linear regression used: male age-dependence from

10-knot thin-plate spline⁴⁴, survey year, menopause status, age, poverty income ratio, smoker status, insurance status, diabetes status, all ethnicities and education.

“

Adjusting for sociodemographics and other variables does not adjust for unknown confounders—even if women are matched for a slew of variables, HRT users are intrinsically different to non users, especially when they are such a rarefied 4% group. This has led to over estimation of “benefits” of HRT in past studies that have been disproven in RCTs. So, suggesting HRT nullifies many of the effects of menopause is over interpreting these data. **Thus I feel the claim that this study shows “Most of the physiological jumps are attenuated by hormone replacement therapy” is unsubstantiated by the data with only 4% using HRT and the fact that over and over again we have seen HRT users are different to non users due to a healthy user bias unexplained by simple sociodemographic characteristics.**

We agree and have now softened our claims throughout. We emphasize that adjusting for sociodemographic variables can not wholly remove healthy user bias. We added the study design to the title “multi-cohort retrospective study” to emphasize its observational nature. We added a limitations section. We use the phrase “HRT *appears* to reduce the jumps”.

Although no statistical test can remove all biases, we added several new salient tests

- 1) Concordance in HRT effect between NHANES (cross-sectional, propensity score matched) and newly analyzed Clalit (longitudinal, matched for clinic, age and birth year)
- 2) A second method for control of NHANES user bias using regression rather than propensity score matching did not change the results.
- 3) Doubling the inclusion of HRT users (from 4% to 9%, including new cohort waves) did not change the results.

I am interested in how the authors reconcile their data with the findings from Shen et al that consistent nonlinear patterns in molecular markers of aging, with substantial dysregulation occurs at two major points, namely at approximately 44 years and 60 years of chronological age. In both sexes. They reported that the transition point in women observed at approximately 55 years of age is not solely attributed to female menopause but, rather, represents a common phenomenon in the aging process of both sexes. (Nature Aging Aug 2024)

We attempted to reconstruct these dysregulation point results using NHANES data and were not able to. There are several issues with the data used by Shen et al including only 3 individuals past age 70. A previous study with a larger dataset and better age coverage, Lehallier et al (2019), identified peaks instead at ages 34, 60 and 78 using proteomics, with no peak at 44; the peak at 60 is small, about 20% of the size of the peak at 78, suggesting at least partial disagreement with Shen et al.

Lehallier, B. et al. Undulating changes in human plasma proteome profiles across the lifespan. Nat. Med. 25, 1843–1850 (2019).

Comments regarding the introduction:

The following statement is not quite correct “This loss of estrogen and other feedback hormones of the hypothalamic-pituitary-ovarian axis causes follicle-stimulating hormone (FSH) and luteinizing hormone (LH) to rise sharply”. The trajectories of both E and FSH around the FMP vary and I suggest this be acknowledged (European Journal of Preventive Cardiology, Volume 23, Issue 7, 1 May 2016, Pages 694–703, <https://doi.org/10.1177/2047487315607044>)

We thank the reviewer and removed this statement, We added the mention of variation to this sentence: “Menopause is due to exhaustion of ovarian follicles, with sizable fluctuations in the years before menopause, ultimately leading to a drop in estrogen and progesterone and rise in FSH, with variability between individuals⁴.”

The way the introduction reports changes related to menopause also implies these changes occur after the final menstrual period (FMP) but the SWAN data has shown that dramatic changes are occurring from the early perimenopause.

We have now removed the incorrect sentence, and noted perimenopausal changes in the sentence mentioned in the previous comment.

I suggest the following needs modification to accurately represent what is known: “the Study of Women's Health Across the Nation (SWAN), showing surges in arterial stiffness within a year of the FMP,” ie endothelial dysfunction. SWAN showed “arterial stiffness” declines from the early perimenopause before estradiol levels have fallen, and the impairment in endothelial function does not accelerate after the FMP as this statement incorrectly suggests (see Moreau KL et al J Clin Endocrinol Metab, 2012 97(12):4692–4700).

We have now removed the sentence about arterial stiffness.

Similarly if increased central fat was just due to estrogen deficiency it should accelerate after the FMP when estradiol is at its lowest, but it accelerates in the menopause transition when women are still making estradiol (Samargandy et al Menopause. 2021 Mar 1;28(6):626–633)

We now remove the detailed list of effects including central fat in the introduction. Instead we say “Because sex hormone receptors are expressed in almost all human tissues, with effects on brain, vasculature, bone, liver, and adipose tissue^{6–8}, these hormonal dynamics reverberate beyond fertility to many physiological systems³.”

I caution the implication that big data is better than a study like SWAN. The strength of SWAN is the impeccable documentation of the actual menopause transition in each participant. This is poorly documented in the big data sets and thus becomes **more about age than menopausal status**.

We now explicitly mention this in the introduction “Such longitudinal studies are crucial due to their exquisite menopause documentation. “ and in the limitation section “While deconvolution sharpens the timeline, it cannot replace the granularity of a prospective longitudinal study with detailed menopausal tracking that follows the same individuals through the transition.”

Finally, in the results:

I am not sure if the authors are referring to their study populations or globally (the ref is global) but the average age of menopause varies from ~47 in India to ~51.5 in Australia so that is not varying by 2 years as suggested.

We now clarify the text. We tested the sensitivity of our results to the shape and position of the FMP age distribution. For this purpose we shifted it by +/-4 years and found the results differed negligibly. We had initially mentioned average age of menopause varies by about +/-2 years, we have now used more clear language.

In Fig S1 we note “The YoGlen method is insensitive to variations in the FMP distribution. Simulated data was a step function at $t=0$ convolved with a normal FMP distribution $N(\mu_0, \sigma_0)$ with $\mu_0=50$ years and $\sigma_0 = 4$ years. a) Original reference FMP distribution (black) and perturbed distributions: one that is shifted by 4 years (green, $\mu = \mu_0 + 4$) (**differences between mean FMP age in different countries span about 4 years¹⁸**) and another that is wider due to added noise (orange, $\sigma^2 = \sigma_0^2 + 4^2$) (larger than typical FMP self-report error noise ~ 1 year^{39,47}).”

Reviewer #2 (Remarks on code availability):

I am not a statistician- I am only able to comment on the data set which I believe is inappropriate for the analysis because of the limitations I have described. This cannot be fixed with clever statistics

We agree that statistics can not replace a prospective study with careful menopause annotation. We now discuss limitations of our study more fully - the revised title includes “retrospective”, we discuss potential biases of observational studies in the discussion, and remark that dedicated prospective menopause studies are essential to test the conclusions. We also add new controls - analyzing HRT in Clalit using a longitudinal design (females matched from the same clinics and same average lab tests before HRT use), doubled the population of HRT users in NHANES, and present the agreement between the two datasets, which we believe makes a stronger argument that the results are generalizable.

Reviewer #3 (Remarks to the Author):

expertise: data deconvolution

This paper examines the effects of menopause on the results of various common lab tests and the likelihood of developing a variety of age-related conditions. Experiments use a substantial (1.3M individuals) data set of lab-test results and diagnostic codes to inform the analysis. The authors deconvolve the trajectory of group-averaged biological function measurements against age with the estimated distribution of age-of-final-menstrual-period (FMP) to estimate the trajectory against time from FMP, which substantially sharpens the menopausal effect and highlights step changes.

We thank the reviewer for this endorsement.

In my view the deconvolution method is reasonable to highlight the effect qualitatively, but does not provide a clear quantitative estimates of menopause-induced change. The conclusions the authors draw largely align with this view, although occasionally they make quantitative statements, such as "Osteoporosis has a ten-fold rise in slope after menopause". I would suggest the deconvolution is not necessary to make the observation that menopause strongly increases the likelihood of osteoporosis, the precise effect of the deconvolution on the slope against age is a little hard to interpret so **precise quantitative statements should be avoided, and perhaps the statement could be informed by data trajectories both before and after the deconvolution to give a more complete view.**

We have now removed precise quantitative statements regarding slopes.

In general the deconvolution idea may be a bit overemphasised in the paper. **Many of the key conclusions about the effect of menopause on biological function and disease prevalence could be supported with the raw data**, although the deconvolution does bring it into sharp relief. Another option would be to **use the male cohort as a baseline against which to compare.**

We now emphasize the male cohort as a baseline for comparison. In Fig 1,2 ... we show the male results using the same algorithm (as if males have a menopause with the same age distribution). These curves do not show jumps in general.

We write in the results "We compared these to males, where we perform the same adjustment as females by deconvolving the same FMP distribution (apparent menopause time $t=0$ for males can be interpreted as age 50). Males did not show jumps around $t=0$, and instead their mean lab tests changed gradually (Fig 2a, blue lines)."

We also write in the discussion "Notably, we find that males lack these abrupt physiological shifts, suggesting that the cliff is a consequence of the menopausal transition rather than generalized aging."

Overall however I **do not have substantive methodological concerns** and believe **this paper makes some useful insights that confirm more anecdotal knowledge** about the strong effect of menopause on biological function and disease prevalence.

We thank the reviewer for this endorsement.

In summary, we thank the reviewers for these comments that helped us add new analyses, clarify the text, cite previous work and address limitations. We believe that the revised manuscript is more clear, rigorous and accurate.

Dear Editor,

There were no specific revisions asked for by the reviewers. We received the comments below.

Sincerely,

Uri Alon

Reviewer #1 (Remarks to the Author):

This study uses a large data set to confirm and expand on expected changes in the menopause transition. The authors have responded appropriately to the reviewers' concerns.

There are no real surprises and no mechanistic data. The cataloging of laboratory changes beyond what is available in SWAN and other studies is worthwhile, but the authors should be careful not to exaggerate the significance.

Reviewer #3 (Remarks to the Author):

I am happy with the changes the authors have made to address the concerns I raised. More broadly they appear to have addressed the concerns of the other reviewers carefully and comprehensively, although I cannot comment on the biological / medical discussions. I have no further comments on this manuscript.